CRISPR-MEDIATED ANTIVIRAL DEFENSE IN *SULFOLOBUS*: A VERSATILE TOOL AND A DANGEROUS WEAPON FOR SETTING UP SAFEGUARDED INDUSTRIAL BIOPROCESSES

Salvatore Fusco

Dottorato in Scienze Biotecnologiche – XXVII ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XXVII ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II

CRISPR-MEDIATED ANTIVIRAL DEFENSE IN *SULFOLOBUS*: A VERSATILE TOOL AND A DANGEROUS WEAPON FOR SETTING UP SAFEGUARDED INDUSTRIAL BIOPROCESSES

Salvatore Fusco

Dottorando:	Salvatore Fusco
Relatore:	Prof.ssa Simonetta Bartolucci
Coordinatore:	Prof. Giovanni Sannia

Ai miei genitori

“Nothing makes sense in biology
except in the light of evolution”

Theodosius Dobzhansky

Contents

Abstract

Since centuries microorganisms have been used to produce fermented products. Nowadays, the consciousness about the importance of these bags to set up eco-friendly industrial workflows is pronounced more than ever. Nevertheless, bio-based processes are much more 'delicate' than chemistry-based counterparts. For instance, microbes are susceptible to infections by viruses that are naturally present in raw matrices, such as the milk for diary industries. Moreover, viruses are the most abundant biological entities on Earth, thus the probability that a give microorganism could encounter such 'predators' is very high. In this regard, huge losses of fermenting biomasses have been already documented, thus leading to open the way to search for new solutions to meet this issue. One of the antiviral mechanisms that could be exploited to obtain resistant strains is the CRISPR-Cas system. This latter relies on the incorporation, upon infection, of sequences from the foreign nucleic acid into the host chromosome at specific CRISPR loci. A locus is an array repeats (similar in length and sequence) interspaced by spacers (dissimilar in sequence) that derive from foreign genetic elements, such as viruses and plasmids. Transcription of these loci generates long precursor pre-crRNAs that are subsequent processed to short crRNAs. Thereafter, these mature RNAs provide specificity to a ribonucleoprotein complex that degrades the invading nucleic acid.

To test the possibility of exploiting the CRISPR-Cas system to produce Virus Insensitive Thermophilic Strains (VITSs), it was resolved to analyse host-virus relationships between the thermophilic archaeon *Sulfolobus solfataricus* and two related member of the *Fuselloviridae* viral family, i.e. SSV1 and SSV2. Intriguingly, although these two viruses show strong similarities, such as virion morphology, genome sequence, gene synteny and expression patterns, it was shown that they are capable of inducing distinct hot reactions. Indeed, whereas SSV1 did not significantly affect the host gene expression, a strong host response was reported upon infection by SSV2. In particular, more than a hundred genes were differentially expressed during the growth of the SSV2-infected strain and, even more interesting, was the up-regulation of both CRISPR loci and *cas* genes. Data reported herein strongly suggest that the host response is influenced by the ability of the virus to control its own gene expression and genome replication during lysogenic growth. In this regard, it was found that SSV1 is a temperate virus that regulates the expression of its genes towards the activity of the transcription repressor F55. As a result, the SSV1 copy number is kept quite low and constant in the lysogenic cells. On the other hand, SSV2 lacks such a regulator and it is incapable of self-regulating its replication; therefore, the infection by SSV2 represents a mush more stressing load for the host. Moreover, this study highlighted an interesting feature of the host-viral interaction in the frame of the CRISPR-response, i.e. host cells containing an integrated copy of the viral genome (provirus) are forced to develop a surviving strategy in order to avoid self-attack by the CRISPR-system. This finding has cast doubt on the feasibility of using this system to challenge viral infections, at least as regards integrative viruses. Further implication and possible solutions to this potential drawback will be herein discussed.

Riassunto

I microrganismi, soprattutto quelli fermentanti, sono utilizzati da secoli per i processi produttivi più svariati. Classici esempi sono quelli della produzione di vino, birra, formaggi e pane. Infatti, si è ben presto compreso il loro impareggiabile valore nel trasformare le materie prime in prodotti con nuove ed attraenti caratteristiche organolettiche. Inoltre, a rendere sempre più richiesto l'impiego di batteri e lieviti nei processi industriali, vi è la capacità di questi ultimi di alterare le proprietà chimiche del prodotto (come il pH) e di produrre molecole con attività antibiotica che facilitano la conservazione dei prodotti fermentati. Il ruolo dei microrganismi nei processi industriali è divenuto ancora più importante grazie a numerose scoperte scientifiche che, nel corso degli ultimi decenni, hanno permesso di modificarli geneticamente rendendoli sempre più indispensabili come '*bio-factories*', non solo per la produzione di prodotti alimentari. Infatti, da tempo si sono consolidate metodiche di produzione di molecole bioattive, quali vaccini, ormoni e anticorpi, nonché di biomasse microbiche da utilizzare come probiotici per complementare la dieta sia dell'uomo che del bestiame. Altrettanto importante è la possibilità di sostituire la sintesi chimica di molecole d'interesse commerciale con quella catalizzata da enzimi prodotti per via ricombinante in ceppi microbici, in modo da rendere sempre più ecosostenibili i processi di produzione industriale del futuro. Ciononostante, l'utilizzo di microrganismi pone numerose problematiche che riguardano, ad esempio, l'insorgenza di ceppi patogeni per l'uomo oppure dei possibili cali della resa produttiva dovuti a un depauperamento della biomassa microbica.

A tal proposito, non è raro che siano riportati casi d'ingenti perdite produttive dovute a infezione a carico della biomassa fermentante da parte di virus o batteriofagi (o fagi), che possono essere presenti naturalmente all'interno delle materie prime utilizzate. Infatti, come la maggior parte degli organismi viventi, i microrganismi sono spesso 'prede' di virus. Poiché è stato dimostrato che questi ultimi sono le entità biologiche più abbondanti sul nostro pianeta, è molto probabile che nel corso della propria esistenza ogni organismo sia infettato almeno una volta. Inoltre, la loro propensione ad accumulare facilmente mutazioni li rende particolarmente adatti a fronteggiare le variazioni ambientali, come ad esempio lo sviluppo d'immunità da parte dell'ospite. Questo è ancora più vero se il sistema virus-ospite è costituito da un fago ed un batterio. Infatti, questi ultimi sono altresì capaci di adattarsi velocemente in risposta agli stimoli ambientali e hanno sviluppato numerose strategie per difendersi dalle infezioni virali. A tal proposito, è stato ipotizzato che il 'conflitto' tra microrganismi e virus abbia rappresentato, e rappresenti ancora, la forza evolutiva più importante all'interno delle comunità microbiche. Questa continua 'corsa agli armamenti' è stata descritta molto efficacemente da Leigh Van Valen nel 1973 utilizzando un celebre romanzo di Lewis Carroll "*Through the Looking-Glass, and What Alice Found There*". La famosa ipotesi, prende il nome di *evolutionary arms race hypothesis* o *Red Queen hypothesis*, e postula che l'adattamento del predatore (il virus) genera una pressione selettiva sulla preda (l'ospite) che a sua volta si ripercuote sul predatore stesso che deve contro-adattarsi per sopravvivere. Come risultato si ottiene che entrambi i protagonisti della 'lotta' vivano in un equilibrio instabile in qui devono continuamente adattarsi per aumentare le loro probabilità di sopravvivenza. Questa ipotesi giustifica largamente le innumerevoli strategie evolute dai microorganismi per difendersi dai virus e di conseguenza le contro-strategie impiegate da questi ultimi per evitare di estinguersi.

L'involucro cellulare (membrana plasmatica e parete) rappresenta la prima barriera alle infezioni virali, infatti, il virus deve essere in grado di adsorbirsi alla superfice cellulare al fine di invadere la cellula ospite. Nella maggior parte dei casi, quest'evento richiede l'interazione tra una proteina di superficie della particella virale e uno specifico recettore esposto dall'ospite. È facile pensare che l'ospite possa sfavorire quest'interazione, ad esempio: i) mutando il recettore, ii) impedendo stericamente l'interazione, producendo una matrice extracellulare oppure iii) secernendo molecole simili al recettore che possano competere, interagendo con la proteina di superfice del virus. Di conseguenza, per infettare l'ospite, i virus possono evolvere: i) nuove varianti delle proteine di superficie in grado di interagire con il recettore mutato dell'ospite, ii) acquisire, attraverso trasferimento genico orizzontale, geni codificanti per enzimi che degradano le componenti della matrice extracellulare oppure iii) evolvere proteine di superficie che interagiscono con altri recettori. Se questa prima linea difensiva fallisce, il virus può iniettare il proprio genoma (DNA o RNA) all'interno della cellula ospite. Ciononostante, l'invasore si trova a dover fronteggiare altre barriere intracellulari che possono riconoscere l'acido nucleico virale come estraneo e degradarlo. Tra i più caratterizzati figurano i sistemi di modifica e restrizione, che si basano sull'espressione da parte dell'ospite di due enzimi: i) uno con attività transferasica (più comunemente una metil-transferasi) che modifica il DNA endogeno e ii) un enzima di restrizione, ovvero, una endonucleasi sequenza-specifica che taglia il DNA esogeno non modificato. L'efficacia difensiva di questi sistemi dipende dalla processività dell'enzima di restrizione, infatti, se il DNA estraneo viene modificato, non sarà più possibile riconoscerlo come 'non-self' e degradarlo. Alcuni virus sfruttano proprio questa 'falla' nel sistema per sfuggire alla restrizione enzimatica, ad esempio, esprimendo a loro volta enzimi di modifica oppure proteine che accelerano la processiva della metil-transferasi endogena. Sebbene questi sistemi abbiano permesso l'evolversi di metodiche d'ingegneria genetica tutt'oggi utilizzate, hanno trovato scarso utilizzo a livello industriale per rafforzare le difese delle biomasse microbiche.

Mentre i meccanismi descritti sinora possono essere paragonati all'immunità innata degli eucarioti, il sistema di difesa antivirale noto con il nome di CRISPR-Cas è sicuramente il progenitore dell'immunità adattativa. Infatti, si tratta di un sistema adattativo ed ereditabile che si basa sulla costruzione di una 'memoria' delle precedenti infezioni, attraverso l'incorporazione, all'interno del genoma dell'ospite, di brevi sequenze di DNA derivanti da virus o plasmidi. In particolare, le sequenze riconosciute sul genoma estraneo '*protospacers'* sono inserite come nuovi '*spacers'* all'interno dei loci CRISPR tra due sequenze identiche chiamate '*repeats'*, in modo da mantenere una classica struttura '*repeat-spacer-repeat*'. All'interno di un singolo locus possono essere presenti da uno fino a un centinaio di unità '*repear-speacer*', dove le '*repeats*' sono identiche in lunghezza e in sequenza all'interno di uno stesso locus, mentre gli '*spacers'* differiscono nella loro sequenza e mostrano spesso identità con DNA virali o plasmidici. Affinché il sistema sia funzionale, un locus CRISPR deve essere trascritto in un lungo precursore noto con il nome di pre-crRNA (*precursor CRISPR RNA*), che sarà in seguito processato in corti crRNAs maturi che potranno essere utilizzati come guida nel riconoscimento dell'acido nucleico estraneo. Il tutto dipende dall'attività di numerose proteine Cas che intervengono nelle tre fasi fondamentali del processo, ovvero, l'adattamento, la trascrizione-maturazione ed l'interferenza. Sulla base delle differenze nel funzionamento delle proteine Cas coinvolte è stato possibile distinguere tre tipi principali di sistemi CRISPR-Cas, che differiscono: i) nella modalità di selezione della sequenza

'protospacer', ii) nel tipo di proteine coinvolte e nel tipo di maturazione del precursore pre-crRNA, iii) nella composizione del complesso ribonucleoproteico che effettua l'interferenza, nonché iv) nel target finale, che può essere per alcuni sistemi il DNA (tipi I e II) mentre per altri sia il DNA che l'RNA (tipo III). La scoperta di un'immunità adattativa nel regno procariotico, che include *Bacteria* e *Archaea*, ha immediatamente aperto la strada a ricerche volte a comprendere come costruire ceppi microbici selettivamente resistenti contro un determinato tipo di virus. D'altronde, per sfruttare al meglio le potenzialità di tali sistemi, è necessario studiare e caratterizzare numerosi sistemi ospite-virus al fine di comprendere i meccanismi regolativi e le funzioni delle proteine coinvolte nel processo.

Nel presente progetto è stato utilizzato come organismo modello, per lo studio dell'interazione ospite-virus, il crenarchaeon termofilo *Sulfolobus solfataricus*. Tale scelta è stata basata su recenti analisi di genomica comparativa che hanno evidenziato la presenza massiva, proprio nei genomi di organismi termofili, di geni deputati a funzioni di difesa contro elementi genetici estranei. Infatti, sono state identificate delle vere e proprie 'isole di difesa' all'interno delle quali, si pensa, possano crearsi numerose opportunità di ricombinazione che potrebbero portare all'evolversi di nuovi sistemi di difesa. Inoltre, l'analisi del genoma di *S. solfataricus* ha mostrato la presenza di ben sei loci CRISPR e numerosi geni *cas*. Diversamente da altri *Archaea* e *Bacteria*, *S. solfataricus* codifica per differenti tipologie di complessi effettori (complessi ribonucleoproteici), ovvero: i) sottotipo IA, anche noto come complesso CASCADE, ii) sottotipo IIIA (complesso CSM) e iii) sottotipo IIIB (complesso CMR). Questi complessi gli permettono di degradare sia il DNA (CASCADE e CSM) che l'RNA (CMR), dunque, rendendolo un organismo d'eccezione per studiare anche l'interazione tra più sistemi CRISPR all'interno dello stesso ospite. Ad aggiungere un ulteriore livello di complessità, vi è la scelta dei virus utilizzati per infettare *S. solfataricus*. I due virus impiegati negli esperimenti descritti di seguito sono membri della famiglia virale *Fuselloviridae*, ovvero, SSV1 e SSV2. *Sulfolobus* spindle-shaped virus 1 (SSV1) è il membro fondatore della famiglia, poiché è stato il primo a essere isolato e caratterizzato. Sebbene il suo ospite naturale sia *Sulfolobus shibatae* B12 isolato da una pozza sulfurea in Beppu (Giappone), SSV1 è in grado d'infettare anche l'ospite permissivo *S. solfataricus* P2 isolato in località Pisciarelli (Pozzuoli). SSV1 è l'unico virus UV-inducibile della famiglia *Fuselloviridae* e pertanto ha offerto un'opportunità unica di studio. Numerose analisi condotte negli ultimi trent'anni hanno permesso di caratterizzare a fondo l'interazione ospite-virus tra SSV1 e i suoi ospiti. In particolare, l'analisi dell'espressione genica di SSV1 ha portato non solo alla definizione della mappa trascrizionale virale, ma ha anche permesso di gettare le basi per comprendere come la trascrizione è regolata negli *Archaea*. Quando SSV1 infetta una cellula ospite, il virus integra una copia del proprio genoma all'interno del cromosoma ospite a livello di un gene codificante per un arginil-tRNA, formando quello che è noto come provirus. D'altra parte, diversamente da altri virus batterici UV-inducibili (come ad esempio il fago lambda), SSV1 è presente nella cellula ospite anche con un numero di 4-5 copie episomali (plasmidiche, non integrate). Questo dato è in accordo con l'identificazione di un ristretto numero di geni espressi dal virus durante la lisogenia che gli assicurano sia una replicazione basale che la produzione di progenie anche in assenza dello stimolo induttivo. Questo è possibile grazie alla natura non litica di SSV1. Appena le cellule lisogeniche sono esposte a luce ultravioletta, si può osservare un '*pattern*' d'espressione virale cronologicamente regolato che comprende l'espressione di: i) un corto trascritto UV-inducibile chiamato T_{ind},

ii) i trascritti precoci (T_5, T_6 e T_9), iii) i trascritti tardivi ($T_{1/2}$, T_3, T_x, and $T_{4/7}$) e iv) quelli estesi tardivamente ($T_{4/7/8}$). Questa cascata di eventi porta, in seguito, all'induzione della replicazione virale e ad un conseguente aumento del numero di copie di genomi virali. Ciononostante, finora è stato possibile solo ipotizzare le funzioni dei geni virali in base ai tempi d'espressione. Infatti, la mancanza di dati strutturali e funzionali, ha impedito qualsiasi ipotesi sul coinvolgimento dei singoli geni di SSV1 nei processi d'ingresso/rilascio delle particelle virali, della replicazione e della trascrizione. Inoltre, i meccanismi regolativi che permettono a SSV1 di mantenere uno stato lisogenico stabile all'interno della cellula ospite non erano ancora noti. Per quanto riguarda invece il *Sulfolobus* spindle-shaped virus 2 (SSV2), erano evidenti alcune similitudini con SSV1, quali l'identità della sequenza genomica (65%), l'organizzazione del genoma stesso e la sintenia genica. Tuttavia, SSV2 è un virus la cui replicazione è indotta dallo stato metabolico/fisiologico dell'ospite, piuttosto che da uno stimolo esterno come la luce ultravioletta. Infatti, almeno nell'ospite naturale *Sulfolobus islandicus REY 15/4*, l'induzione della replicazione di SSV2 è regolata dall'ingresso nella fase stazionaria della crescita dell'ospite. Fenomeno, quest'ultimo, che non avviene nell'ospite permissivo *S. solfataricus*, nel quale SSV2 mostra un tasso di replicazione paragonabile a quello che si raggiunge nell'ospite naturale solo in fase stazionaria. Questo ha fatto ipotizzare che un regolatore prodotto da *S. islandicus* (e non da *S. solfataricus*) potesse essere responsabile delle differenze osservate. Le discrepanze nella modalità d'induzione della replicazione sono ancora più sorprendenti alla luce di un recente studio in cui è stato riportato che la mappa trascrizionale di SSV2 è praticamente sovrapponibile a quella di SSV1. Partendo dalle differenze e dalle similitudini appena descritte, in questo progetto si è proposto di studiare: i) le basi molecolari responsabili della regolazione dell'espressione genica e dell'induzione della replicazione di SSV1 ii) la risposta dell'ospite *S. solfataricus* in seguito all'infezione dei due virus e iii) l'interazione ternaria virus-ospite-virus in un ceppo infettato contemporaneamente da SSV1 e SSV2.

Sebbene SSV1 abbia rappresentato negli ultimi trent'anni un modello per studiare l'interazione ospite-virus nel dominio *Archaea*, i meccanismi alla base dello 'switch' dalla fase lisogenica all'induzione della replicazione virale erano ancora poco chiari. Un iniziale passo avanti nella comprensione di questo processo è stato fatto grazie ad esperimenti pilota condotti durante la tesi magistrale, nell'ambito di un progetto Erasmus (2009/2010), sotto la supervisione del professor Qunxin She presso il Danish Archaea Centre dell'Università di Copenaghen (Danimarca). Infatti, analisi preliminari di microarray su cellule di *S. solfataricus* infettate da SSV1, hanno permesso l'individuazione di un nuovo gene virale (*f55*) espresso da SSV1 in cellule lisogeniche. In particolare, la localizzazione di *f55* lasciava pensare che la proteina da esso codificata (F55) potesse svolgere un ruolo chiave nel mantenimento della lisogenia. Infatti, *f55* è localizzato all'interno di una regione del genoma virale che ricorda quella che codifica per il repressore CI nel genoma del fago UV-inducibile lambda. Inoltre, predizioni di struttura secondaria e terziaria indicavano che la proteina F55 potesse assumere una strutturazione di tipo Ribbon-Helix-Helix (RHH), tipica di repressori della trascrizione. Ancora più interessante e stata l'individuazione dei classici segnali di legame al DNA riconosciuti da questi regolatori, cioè, sequenze ripetutene in tandem all'interno dei promotori dei geni espressi in risposta all'irradiazione con luce ultravioletta. Tutte queste evidenze lasciavano ipotizzare un ruolo di F55 nel regolare l'espressione di questi geni importanti per l'induzione. Per questi motivi, il putativo repressore trascrizionale F55 è stato oggetto di caratterizzazioni *in vitro* e *in vivo* nell'ambito di questo progetto di dottorato, al fine di

comprendere il suo ruolo nella regolazione del ciclo vitale di SSV1 e come questo potesse ripercuotersi sulla risposta dell'ospite. Gli esperimenti riportati confermano la capacità di F55 di interagire *in vitro* con numerose sequenze operatrici ricavate dai promotori dei principali geni espressi durante la transizione dallo stato lisogenico a quello indotto. Inoltre, la messa a punto di un protocollo d'irradiazione standardizzato, ha consentito di studiare l'attività regolatrice di F55 *in vivo*, confermando il suo ruolo di repressore trascrizionale. I dati prodotti, non solo forniscono il primo esempio di fattore trascrizionale codificato da un fusellovirus la cui funzione regolatrice è stata sperimentalmente dimostrata, ma confermano, inoltre, che SSV1 è un virus temperato in grado di autoregolare la propria espressione genica e di conseguenza d'instaurare una pacifica co-esistenza con la cellula ospite, che non ha portato all'attivazione del sistema CRISPR-Cas da parte dell'ospite.

Tutt'altro scenario è fornito da SSV2, infatti, i dati discussi di seguito dimostrano che l'interazione da parte di questo fusellovirus comporta l'attivazione trascrizionale non solo dei loci CRISPR di *S. solfataricus*, ma anche di geni *cas* codificanti per proteine che costituiscono il noto complesso d'interferenza CASCADE. Questa evidenza è in totale contrasto con il pattern d'espressione di SSV2 in cellule lisogeniche, infatti, solo minime differenze sono state riscontrate quando paragonato con quello di SSV1. Esempi sono: i) l'espressione esclusiva di *f55* da parte di SSV1, che non presenta un omologo nel genoma di SSV2 e ii) l'espressione di *d79* e *b223* di SSV2, i cui omologhi non sono espressi da SSV1. Mentre la presenza del regolatore F55 conferma che il virus SSV1 sta controllando autonomamente la fase lisogenica, mantenendo anche un basso numero di copie, l'espressione della DnaA-like B223 da parte di SSV2 lascia immaginare un tentativo da parte di ques'ultimo di far convergere il macchinario replicativo dell'ospite verso il proprio genoma, dunque, mantenendo un tasso di replicazione più elevato nella fase lisogenica. Ancora più interessante sono i dati ottenuto a seguito di una selezione di cloni singoli dal ceppo di *S. solfataricus* infettato con SSV2. In particolare, è stato dimostrato che l'attivazione del sistema CRISPR-Cas ha determinato un'effettiva diminuzione del numero di copie del genoma di SSV2 all'interno dell'ospite, ma purtroppo seppure alcuni cloni con una quantità minima di DNA virale siano stati isolati, non è stato però possibile identificare ceppi definitivamente 'curati' dall'infezione. Quando poi il sito d'integrazione virale all'interno dell'ospite è stato analizzato mediante esperimenti di PCR, è stato subito chiaro che l'impossibilità nell'isolare cloni curati dipendesse da fenomeni d'intrappolamento del genoma virale all'interno del cromosoma dell'ospite. Infatti, i cloni che mostravano una quantità minima di DNA virale, erano gli stessi che mostravano un grado di occupazione del sito di integrazione pari quasi al 100%. L'intrappolamento del provirus all'interno del genoma ospite può essere spiegato prendendo in considerazione che gli unici genomi virali ad esprimere una copia funzionale del gene codificante l'integrasi virale, deputata ad integrare ed escindere il DNA virale, sono quelle episomali, poiché la ricombinazione del DNA virale con quello dell'ospite porta alla 'rottura' di questo gene sulla copia provirale. Se si considera che l'attivazione del sistema CRISPR-Cas dell'ospite porta alla degradazione della maggior parte delle copie virali episomali, è semplice immaginare che questo evento abbia portato ai fenomeni di intrappolamento osservati. Ancora più interessante è la ripercussione sull'ospite di tale circostanza, infatti, analizzando i loci CRISPR delle cellule che mostravano il provirus intrappolato nel loro genoma, si è sorprendentemente osservato che uno dei loci (il locus F) avesse subito delezione di alcuni 'spacers'. L'analisi della sequenza degli 'spacers' deleti ha evidenziato che almeno due di questi fungono da guida nell'interferenza contro il virus SSV2. Questo

dato pone una domanda molto interessante, ovvero: perché un ospite che sta reagendo contro l'infezione virale tende a perdere 'spacers' contro il virus invasore? Questa domanda ha ricevuto una risposta alla luce di alcuni recenti studi che hanno dimostrato che la presenza nei loci CRISPR di 'spacers' diretti contro sequenze endogene non fosse da considerarsi come un'ulteriore evoluzione del sistema verso la regolazione dell'espressione genica endogena (tipo interferenze eucariotico), ma che l'incorporazione di sequenze 'self' fosse in realtà un errore seguito spesso dalla perdita della funzionalità dell'intero sistema. Come correlare queste informazioni con i risultati ottenuti? La risposta è fornita proprio dall'abilità dei fusellovirus d'integrarsi all'interno del genoma ospite, dunque, diventano vere e proprie sequenze 'self'. Infatti, poiché le sequenze delle copie episomali e quella del provirus sono identiche, se il sistema di difesa è in grado di riconoscere e degradare le copie episomali, questo deve poter verificarsi anche con il provirus. Il taglio endonucleolitico a livello del provirus espone inevitabilmente estremità del DNA degradabili da esonucleasi, evento che con molta probabilità coinvolgerebbe anche il cromosoma dell'ospite. A tal proposito, è stato proposto che la pressione selettiva generata da questi eventi d'autoimmunità, abbia favorito l'insorgenza nella popolazione microbica di cloni cellulari con difetti nel funzionamento nel sistema di difesa, come ad esempio, la perdita di 'spacers' utilizzati per contrastare l'infezione di SSV2 dal locus F. Ciononostante, non è possibile escludere il verificarsi di altre possibili mutazioni che abbiano potuto inattivare il sistema antivirale, in modo da evitare fenomeni deleteri per la cellula ospite.

È chiaro che queste evidenze sperimentali possono ripercuotersi negativamente sulla possibilità di applicare tali sistemi per la costruzione di ceppi resistenti a specifiche infezioni virali, soprattutto per quanto riguarda le infezioni a carico di virus integrativi, che potrebbero sfuggire al controllo similmente a quanto mostrato per SSV2. A tal proposito, è stato proposto che una valida alternativa, per aggirare tali fenomeni di autoimmunità, potrebbe essere quella di sfruttare maggiormente sistemi CRISPR-Cas di sottotipo III-B che hanno come target finale l'RNA piuttosto che il DNA. Ad ogni modo, questo limiterebbe l'applicabilità ai soli ospiti che esprimono questo sottotipo di sistema. Un'altra soluzione potrebbe essere quella di utilizzare ceppi con siti d'integrazione mutati, anche se questo implica una conoscenza approfondita dell'interazione ospite-virus. Sebbene lo svantaggio dell'insorgenza dell'autoimmunità sia stato evidenziato da questo studio, i dati riportati sono molto interessanti da un punto di vista evolutivo. Infatti, come anticipato, l'adattamento di una delle parti in gioco (il virus) ha determinato l'instaurarsi di una pressione selettiva sull'altra parte (l'ospite), dunque confermando che: "la corsa evolutiva agli armamenti è una storia infinita!".

Chapter I

– Introduction –

1.1 The host-virus arms race: a never-ending story

Thanks to their amazing capability of adaptation to a variety of environmental conditions, prokaryotes (i.e. *Bacteria* and *Archaea*) are the most abundant organisms on Earth. Nevertheless, it has been predicted that in several niches, bacteriophages (infecting *Bacteria* and referred to hereafter as phages) and viruses (infecting *Archaea*) outnumber their hosts by approximately tenfold [1, 2], thus being formidable predators. Arguably, the fight between prey and predator is the oldest conflict acting on our planet and represents the most dominant and persistent evolutionary driving force in microbial communities [1, 2]. This struggle has led to a never-ending arms race, which has requested to either evolve or acquire (through lateral gene transfer) an extraordinary toolbox of defence mechanisms acting at multiple levels against phage infection [3]. In this regard, it is worth of note that in the 1973 Leigh Van Valen proposed the **evolutionary arms race hypothesis**, also known as **Red Queen hypothesis**, which proposes that environmental competition leads to the establishment of a selective pressure acting on the host and, in turn, to counter-adaptations on the parasite side [4]. This theory explains the competition between hosts and parasites by referring to a famous Lewis Carroll's fantasy novel entitled: "*Through the Looking-Glass, and What Alice Found There*". In this story, the Red Queen tells to Alice:

*"Now, here, you see, it takes **all the running you can do**, to keep in the same place. If you want to get somewhere else, you must **run at least twice as fast as that**!"*

The original hypothesis posits that in continuous co-evolutionary relationships, adaptation on one side (i.e., *all the running you can do*) would lead to the prospective extinction of the other side, unless this latter tries to maintain its fitness through counter-adaptation (i.e., *run at least twice as fast as that*). As a result, both sides live in a precarious balance where they have to continuously evolve in order to keep almost the same fitness level. This evolutionary trend is even more pronounced if the host and the parasite are a microorganism and a phage, respectively. In fact, the fast-forwarding evolution of phages is a selective pressure that constantly acts on microbial communities, thus guiding in turn their evolution [2].

Generally speaking, the first step of the infection is the **adsorption** of a viral particle onto the host cell surface. This is a very complex event because a phage has to recognize a host-specific cell component (such as a receptor) among an uncountable number of molecules forming the host membrane and cell wall. Moreover, this process is made even more intricate because of the array of barriers evolved by prokaryotes to become immune to the infection. There are at least three major categories of **adsorption-blocking mechanisms**, which are: i) mutation or blocking of the phage receptor, ii) production of an extracellular matrix and iii) secretion of competitive inhibitors [1]. To interfere with the phage propagation, the host can modify the structure of the relative receptor or can produce proteins that mask it through stable interaction. Moreover, by producing structured extracellular polymers, the host cell can physically block the interaction between the phage and the receptor. For example, molecules having a role either in the communication with other member of the microbial population (*quorum sensing*) or in the competition with other strains (*antimicrobial molecules*) can interact with specific receptors, thus hindering phage adsorption [1]. As anticipated above, phages have to develop counterstrategies in order to cope with these host traits. In many cases, it has been

demonstrated that phages have evolved new variants of their receptor-binding proteins either for adapting to the mutated receptor or for acquiring novel receptor tropism [3]. Tactics of **digging for receptors** have also been reported in literature. For instance, when the host receptor is hidden by components of the cellular envelope, phages can get access to them by degrading the extracellular matrix thanks to depolymerizing enzymes carried on the tip of the capsid. In some cases, the expression of a host-specific receptor is regulated in response to internal or external stimuli, thus rendering the phage adsorption a much rare event. Indeed, to face this stochastic expression and consequently to increase the chances of infecting the host cell, phages can expand their binding specificity through mutations of genes encoding receptor-binding proteins, to achieve **stochastic recognition of variable host receptors** [3]. When a viral particle (referred to hereafter as virion) stably attaches to the host surface, the viral nucleic acid has to be injected into the host cell. In this regard, it has been shown that this process can be a target of many **superinfection exclusion systems** that use membrane- or cell wall-anchored proteins to prevent phage DNA entry. Intriguingly, genes encoding for such proteins are usually found in provirus (i.e. integrated isoforms of the phage DNA, a.k.a pro-phage). This suggests their importance in phage-phage rather than in host-phage interactions [1].

Once the viral genome has managed to enter into the host cell, still it has to face with a multitude of intracellular antiviral barriers. Three major strategies have been evolved to challenge the establishment of foreign genetic elements (such as viruses and plasmids). One of the first being discovered and the most characterized is the **restriction-modification system**, which is present in 90% of bacterial and archaeal genome sequenced; indeed, it is almost a ubiquitous strategy to discriminate **self** from **non-self** DNA. This system relies on two distinct activities: i) a **modification enzyme** that labels the host DNA, usually through methylation of specific bases within a recognition sequence and ii) a **restriction enzyme**, which is a sequence-specific endonuclease acting only on unmodified DNA. Upon injection of an unmethylated phage genome, the restriction enzyme rapidly degrades the invader DNA. On the other hand, phages have evolved both passive and active mechanisms to avoid restriction-modification systems. It is easy to conceive that, for a foreign DNA, the chances of being restricted depend on the balance between the activities of the modifying and restriction enzymes. If the DNA methyltransferase acts rapidly by modifying the phage DNA before it is recognized by the restriction enzyme, the invader is protected. Moreover, phages with few or none restriction sites in their genome can take over as a consequence of the selective pressure. Alternatively, phages have evolved active counter-mechanisms to accelerate genome protection, such as acquiring genes encoding for: i) proteins that mask the viral DNA, ii) DNA-modification enzymes or ii) proteins stimulating the activity of the endogenous modification enzymes [3].

An astonishing system that prokaryotes evolved throughout the arms race against phages/viruses is the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins) defence system [5]. If other systems can be compared to the **innate immune response** of eukaryotes, this latter might be considered as the ancestor of the **adaptive immune response**. Indeed, the CRISPR-Cas system is a highly adaptive and heritable resistance mechanism that incorporates sequences derived from foreign DNA elements into a small-RNA-based repertoire [5]. Briefly, it consists of a multistep process by which specific small fragments from the foreign nucleic acids are, at first, recognized as being non-self

and then incorporated into the host genome. Subsequently, these fragments (referred as spacers), in conjunction with host Cas proteins, are used as a surveillance and adaptive immune machinery, by means of which incoming foreign nucleic acids are recognized and destroyed or possibly silenced (for further details see par. 1.2). Once again, phages/viruses can counteract by: i) acquiring mutations that hinder the recognition of the target sequence, ii) expressing anti-CRISPR genes that have been suggested to encode for proteins interfering with the formation of the endogenous CRISPR-Cas surveillance complex and iii) encoding for CRISPR-Cas systems able to inactivate the host antiviral defence system [3]. If all these above-mentioned protection mechanisms fail, the last chance of the host cell is the **abortive infection**, which is a programmed cell death activated once the host functions have definitively been corrupted by the phage. Although this "suicide" may appear as nonsense at single cell level, it prevents the spreading of infectious particles thus conferring an indisputable advantage to the surrounding microbial community.

Noteworthy, comparative genomic studies have recently revealed a high abundance of defence systems in the genomes of thermophiles (especially hyperthermophiles) if compared with mesophiles and psychrophiles [6]. Moreover, it has been observed that this group of defence genes clusters into genomic regions known as **defence islands**, i.e.: "strings of continuous genes, at least one of which belongs to a known defence gene families, and are flanked by house-keeping genes" [6]. Arguably, these islands are a "playground" for functional shuffling that can readily lead to the evolution of new defence systems that still remain to be discovered and exploited for biotechnological applications. In this regard, it is notorious that viral infections interfere with any industrial process relying on microbial metabolism for the production either of biocatalysts or of fermented products [7]. As a result, there is a growing interest in developing suitable protocols for the selection of phage-resistant microorganisms. In this context, the CRISPR-Cas system represents a powerful tool to set up safeguarded bioprocesses in the near future.

1.2 CRISPR-Cas system: an overview

1.2.1 Discovery and structural features

Long before its biological role was unravelled, in the 1987 Ishino and co-workers made the first report on the CRISPR-Cas system. These authors found that immediately downstream the gene encoding for isozyme-converting alkaline phosphatase (*iap*) of *Escherichia coli*, there was a set of 29-nt repeats interspaced by unrelated and non-repetitive short sequences (spacers) [8]. Fast-forward 13 years, and the availability of sequenced microbial genomes allowed genome-wide computational searches that led Mojica and colleagues to define this sequences as a family of repeats recurring in many species of *Bacteria* and *Archaea* [9]. Later on, Jansen and co-worker coined the term **CRISPR** (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats) to reflect their characteristic structure [10]. Indeed, a typical CRISPR locus is an array of short direct **repeats** interspersed with **spacer** sequences (Fig. 1). Within a given locus, repeats are identical in length and sequence (Fig. 1, grey boxes) whilst spacers are highly variable sequences though uniform in length (Fig. 1, coloured boxes).

Repeat-spacer units range from as few as one to as many as several hundred, with an average of twenty. Soon after, computational analyses of spacers

revealed that they match sequences from extrachromosomal elements, such as plasmids and viruses. Moreover, other studies found a correlation between the sensitivity to the infection by a specific virus and the absence of spacers matching the genome of that particular virus, thus suggesting for CRISPR loci a plausible function as defence system [8]. Subsequently, by studying the lactic-acid bacterium *Streptococcus thermophilus*, it was finally validated that the CRISPR-Cas system plays a role in challenging foreign genetic elements. In particular, different spectra of phage sensitivity were detected for strains of *S. thermophilus* showing CRISPR loci with different number and type of spacers. Additionally, it was shown that the acquired resistance could be lost if mutations occur either in the spacer or in the complementary sequence within the targeted phage DNA, i.e. the **protospacer** [7].

Sequence analysis of genomes containing multiple CRISPR loci uncovered an additional conserved region, namely the **leader sequence,** which precedes the repeat-spacer array (Fig. 1, black box). This conserved element is an AT-rich region with intra-species but not inter-species conservation. Indeed, it is up to 80% identical within a given genome but it is quite dissimilar among different species. The leader extends over several hundred base pairs, it lacks coding potential and is always found at one side of the CRISPR in a fixed orientation [11]. Recent studies have assigned to the leader sequence a role in guiding the expression of the downstream CRISPR locus as well as in the acquisition of new spacers, a process known as **adaptation**.

CRISPR-associated (*cas*) genes, a set of conserved protein-coding genes associated with the loci, are usually located at one side of the repeat cluster (Fig. 1, coloured arrows). Extensive bioinformatics analyses have shown that the genomes of various CRISPR-containing microorganisms encode approximately 65 distinct sets of orthologous Cas proteins with nucleases (RNase and/or DNase activity), helicases, polymerases and RNA-binding activities. Unfortunately, the nomenclature of these genes depends on the classification criteria used and can easily create confusion [12].

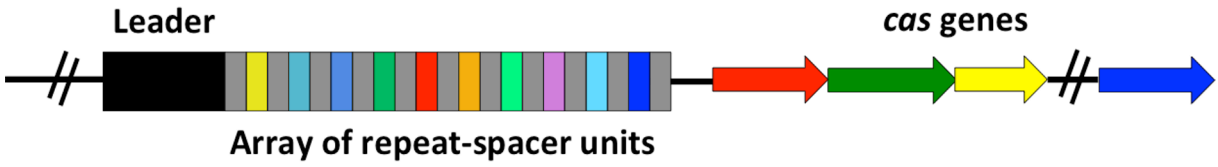


Figure 1. Structure of a CRISPR-Cas locus. Repeat sequences (grey boxes) are interspaced by spacers (coloured boxes) of approximately the same size. The number of repeat-spacer units varies greatly among different microorganisms. An AT-rich leader sequence (black box) is several hundred base pairs long and located at one side of the array. CRISPR-associated *cas* genes are usually located at one side of the locus.

1.2.2 Mechanism of action

CRISPR activity can be ideally divided into three steps: **adaptation**, **expression** and **interference**.

Adaptation: incorporation of new spacers

An essential prerequisite of any adaptive immune system is the ability of acquiring a memory of past invaders in order to face future infections. Therefore, the first step of CRISPR-Cas defence systems consists in the recognition of short DNA sequences on foreign genetic elements (i.e. **protospacers**) and their incorporation into a CRISPR array, as new **spacers** (Fig. 2) [13]. This event will provide resistance (immunization) against foreign genetic elements carrying the protospacer. With few exceptions, new spacers are inserted at the leader-proximal region of the CRISPR locus, with most integrations occurring at the first position in the cluster. The integration is coupled with the duplication of a repeat, thus allowing the maintenance of the typical pattern of spacer-repeat units (Fig. 2). Nevertheless, the loss of one or more repeat-spacer unites has also been observed and may be part of control mechanisms evolved to avoid the continuous growth of the locus [11]. Although our current knowledge of the adaptation process is still in its infancy, a crucial role has been assigned to the universally conserved Cas1 and Cas2 proteins (Fig. 2) [13]. However, several conserved *cas* genes as well as accessory factors might be involved in spacer acquisition [14]. As concerns the selection of the protospacer it has been experimentally demonstrated that, in some cases, it is not a random event but it depends on the presence of a sequence termed **Protospacer Adjacent Motifs** (**PAMs**), i.e., a 2-3 nucleotides motif located at one end of the protospacer sequence [14].

Transcription and processing

Once a spacer is inserted in a CRISPR locus, it is essential to transcribe the repeat-spacer array to generate a long precursor, i.e. a **pre-crRNA** (precursor CRISPR RNA). In general, transcription of the CRISPR loci is constitutive and unidirectional, even if exceptions have been reported in *Sulfolobus spp*. [12]. Additionally, experiments in *T. thermophilus* suggested that the CRISPR-Cas pathway is activated under carbon source limitation, thus allowing speculations about mechanisms of transcription regulation depending on the host metabolism [11]. The initial processing of the pre-crRNA involves endoribonucleolytic cleavages within the repeats to release small **crRNAs** (Fig. 2). Further trimming of these short RNAs leads to the production of mature crRNAs that sequence-specifically guide the interference machinery towards invasive targets. Several modalities of maturation and targeting have been described, leading to the classification of CRISPR systems in three main types, i.e. **type I, II and III** (for further details see par. 1.2.3).

The interference phase

Once generated, mature crRNAs form **crRNA-Cas ribonucleoprotein complexes** (crRNP) and guide, through base-pairing, the recognition of the invasive genetic element (Fig. 2). Several lines of evidence indicated that, in numerous species, the CRISPR machinery recognizes DNA rather than RNA targets, though in *P. furiosus* these systems showed to be able to target RNA [13, 14]. Efficient immune mechanisms must distinguish **self** from **non-self** to avoid autoimmunity and CRISPR-Cas systems are not an exception. In this case, discrimination occurs during

the interference phase. The spacer region of a mature crRNA, which is embedded in a CRISPR-Cas surveillance complex, is not sufficient for such discrimination since it is also complementary to its template in the CRISPR locus on the host chromosome. Nevertheless, the aforementioned PAM motif has been demonstrated to play a relevant role in this process. This short sequence is located at one end of the protospacer sequence (on the invasive genetic element) but is no longer present in the spacer sequence at level of the CRISPR locus. Arguably, the partial mismatch between the crRNA and the protospacer causes a conformational change that leads to the cleavage of the target, whereas the full complementarity between the crRNA and the spacer of the endogenous CRISPR locus spares the host chromosome from interference [15].

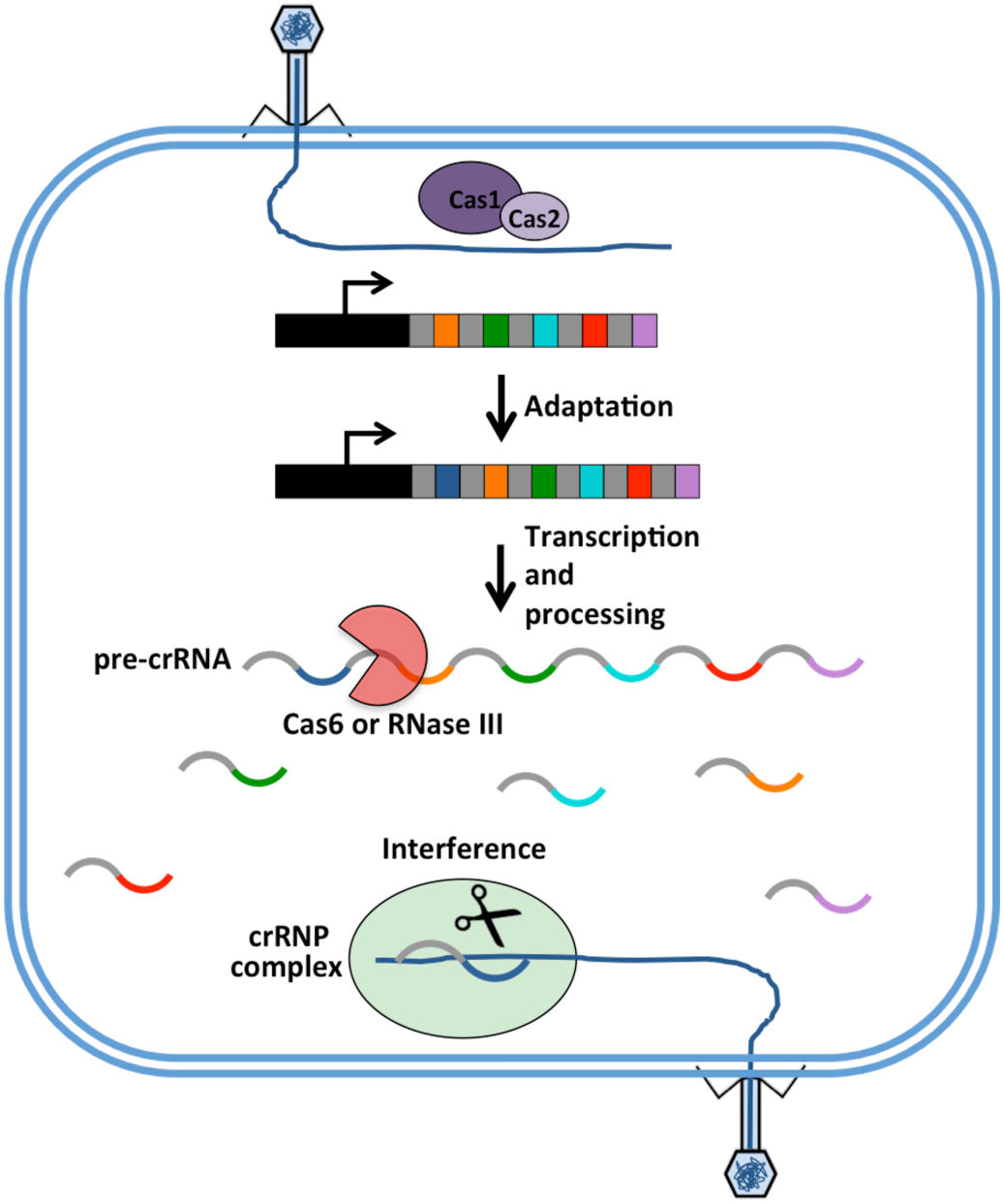


Figure 2. Overview of the CRISPR-Cas mechanism of action. Upon infection, spacers derived from the invading DNA are incorporated into a CRISPR locus (adaptation). The repeat-spacer array is then transcribed into a pre-crRNA, which is processed to short crRNAs (transcription and processing). Upon subsequent infections, mature crRNAs provide sequence-specificity to the ribonucleoprotein complex, which detects and cleaves the foreign nucleic acid.

1.2.3 A new classification

An initial classification, based on the analysis of 40 bacterial and archaeal genomes, divided the CRISPR loci into twelve classes and named *cas* genes using a simple three-letters code. For example, to an *E. coli* cas gene could be assigned the *cse* code (CRISPR system of *E. coli* gene), where a number indicates its position within a *cas* gene cassette (e.g., cse1, cse2) [12]. Although this system offered a simple classification code, it did not take into account the phylogenetic relationship among *cas* genes. Recently, a new polythetic classification was proposed to integrate CRISPR loci and *cas* genes phylogenies and introduced three major types of CRISPR systems (i.e., type I, II, and III) that can be further divided into subtypes (i.e., **type I-(A-F)**, **type II-(A-B)**, and **type III-(A-B)**) [16].

Type I CRISPR–Cas systems

A hallmark of all type I CRISPR-Cas systems is the *cas3* gene, which encodes for a protein with helicase and phosphohydrolase activities. In some subtypes (A, B, and D), separate genes encode for the nuclease and helicase domains of Cas3. To protect the host from the infection, Cas3 interacts with a crRNA-guided surveillance complex formed by other **Cas** proteins as well as by **RAMP** proteins (Repeat Associated Mysterious Proteins) (Fig. 3). The first CRISPR-associated complex being isolated and characterized is that of *E. coli* K12 (type I-E), which was termed **CASCADE complex** (CRISPR-associated complex for antiviral defence). The crystal structure of Cas7 (one of its component) revealed that this protein assembles along the ribose-phosphate backbone of the mature crRNA and keeps it in an under-wound and stretched conformation. The CASCADE complex, recently isolated and studied also in *Sulfolobus solfataricus* (type I-A), cleaves long pre-crRNA into mature 61-nt crRNAs. In the interference phase, the recognition of the target sequence by the ribonucleoprotein complex leads to conformational changes that, in turn, allow the recruitment of Cas3 to cleave the invader DNA (Fig. 3) [12, 16].

Type II CRISPR–Cas systems

To date, CRISPR-Cas systems belonging to the type II have been found only in *Bacteria*. The *cas* gene cluster generally consists of four genes, i.e., *cas1*, *cas2*, *cas9* and either *csn2* (type II-A) or *cas4* (type II-B). Likewise the type I, the final target of the interference complex is the DNA. The multifunctional protein Cas9 has been shown to play a role in both mature crRNAs biogenesis and in the degradation of invading DNA (Fig. 3). Nonetheless, crRNAs maturation for the type II systems requires also: i) a **tracrRNA** (trans-encoded small RNA), which is complementary to the repeat sequences of the pre-crRNA and ii) an endogenous **RNase III** enzyme (Fig. 3). Briefly, Cas9 binds to a pre-crRNA and recruits the tracrRNA that forms a local dsRNA (double-stranded RNA). This is followed by the endonucleolytic cleavage catalyzed by the RNase III. A further maturation might be carried out by Cas3, which trims the crRNA at a fixed distance within the spacer sequence. At last, the mature crRNA-Cas9 complex can detect and degrade the exogenous DNA (Fig. 3) [16].

Type III CRISPR–Cas systems

This kind of systems are widespread in *Archaea* and, unlike those of types I and II, involve both Cas and RAMP proteins. The first step is the binding of Cas6 (a CRISPR-specific endoribonuclease) to the pre-crRNA that is cleaved at the 3'-end of

the repeat (Fig. 3). The crRNA is likely transferred to a distinct complex, i.e.: i) to the **Csm complex** for the subtype III-A (also known as **CASS6**) or ii) to the **Crm complex** for the type III-B (also known as **RAMP module**). In subtype III-B systems, the 3′-end of the crRNA is further trimmed within the spacer sequence (Fig. 3). Whereas for type I and II complexes the final target is the DNA, a type III complex detects and cleaves either DNA (subtype III-A systems) or RNA (subtype III-B systems) (Fig. 3). Type III gene cassettes may lack *cas1* and *cas2* genes, but in all these cases at least an additional CRISPR locus (of either type I or type II) is present and probably provides in trans Cas1 and Cas2 functions [16].

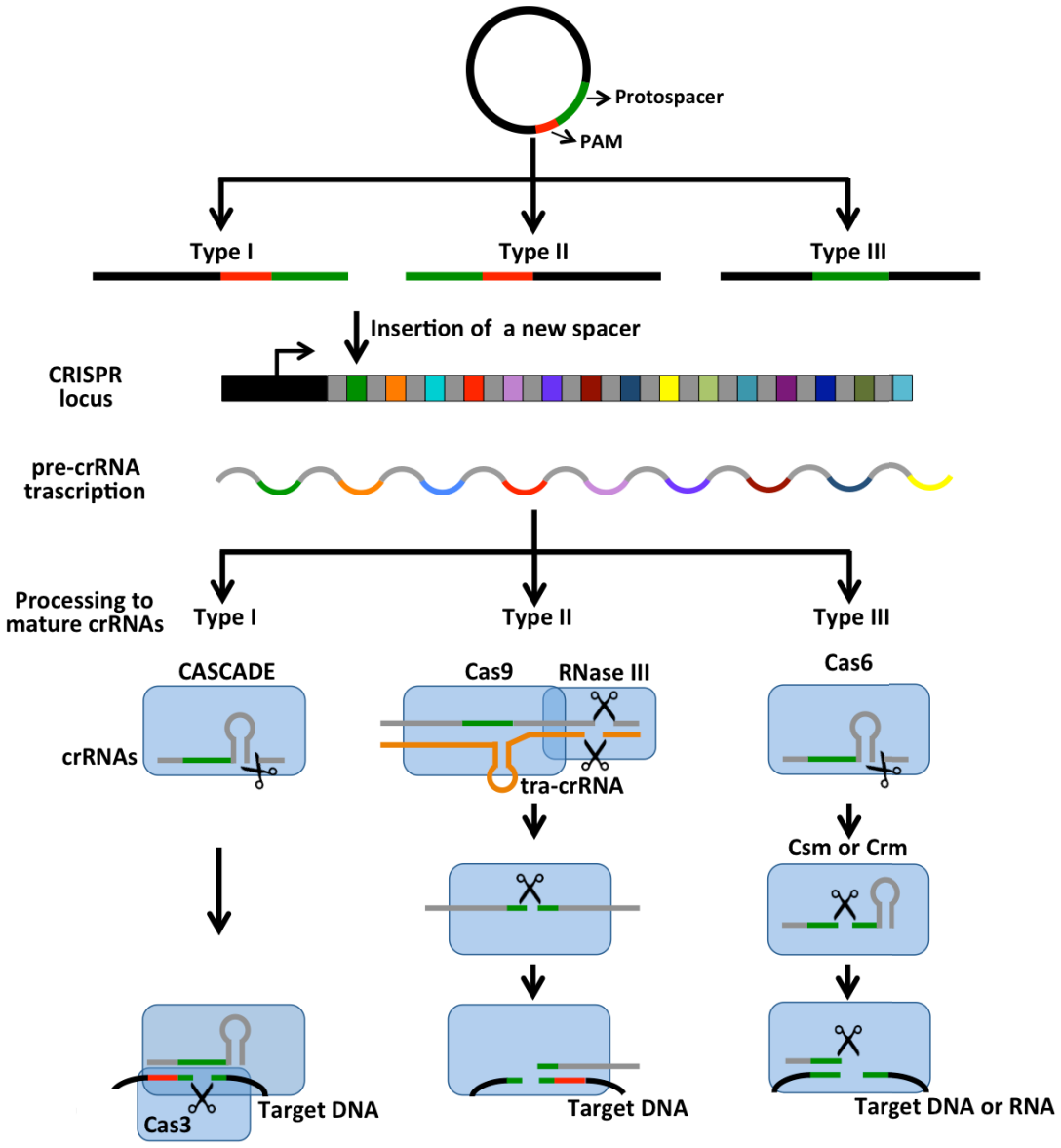


Figure 3. Three major types of CRISPR-Cas systems. With the exception of the type III, the selection of a protospacer from the invading nucleic acid depends on the presence of a PAM motif. During the expression stage, the CRISPR locus is transcribed into a long primary CRISPR transcript (the precrRNA). In type I systems, the CASCADE complex recognizes the pre-crRNA and cleaves it by means of Cas6. The resulting short crRNAs is used to detect and degrade the invading DNA. Instead, type II systems use a trans-encoded small RNA (tracrRNA) and the housekeeping RNase III to bind and process the pre-crRNA into mature crRNA, which are further trimmed by Cas9 within the spacer. In type III systems, although Cas6 is responsible for the initial processing, the crRNAs seem to be transferred to a distinct Cas complex (Csm or Crm).

1.2.4 Self-targeting spacers: dangerous weapons?

The finding that CRISPR spacers mainly match to sequences from foreign genetic elements has directed early experimentations to the elucidation of its role as defence system. However, a recent analysis carried out by Stern and colleagues have shown that self-targeting spacers occur in 18% of all CRISPR-bearing microorganisms [17]. Although this evidence could lead to hypothesize that CRISPR systems participate to unexplored gene regulation mechanisms [18], Stern's results point out that the incorporation of self-targeting spacers leads to autoimmunity and, in turn, to the loss of CRISPR function [17]. Indeed, several lines of evidences cast doubts on the hypothesis that the self-targeting spacers serve to regulate gene expression, i.e.: i) self-targeting spacers do not confer a selective advantage, indeed, they are not conserved over a phylogenetic lineage; ii) many of them have been found at the leader-proximal region of the CRISPR loci, thus indicating that they 'survive' only over a short time window because they are detrimental for the host; moreover, iii) acquisition of self-targeting spacers are frequently associate with the partial or complete loss of CRISPR-Cas functionality [17]. Accordingly, it has been demonstrated that when *S. solfataricus* is transformed with a plasmid that carries an essential gene and is also a target of the CRISPR system, surviving transformants show mutations that eliminate the plasmid-targeting spacer [19]. Intriguingly, these evidences open the way to hypothesize that the CRISPR system might be dangerous for the host itself [20].

1.2.5 Current and prospective applications

Spoligotyping: a spacer-oligotyping method

In the adaptation stage new spacers are incorporated into the CRISPR loci, thus explaining rapid changes in their structure. As a result, within a microbial community, several subpopulations differ in the content and type of spacers. This observation led Kamerbeek and colleagues to develop a spacer-based method for the genotyping of *Mycobacterium tuberculosis* strains, also known as **spoligotyping**. This method is based on a DNA hybridization approach that allows discriminating different strains by means of their spacer repertoires [8]. Later on, spoligotyping has become a consolidated procedure for studying the evolution of bacterial populations and species in the context of genome sequencing projects. Nevertheless, its applicability has to be evaluated case-by-case, because of the high dynamism of CRISPR loci [21].

Engineered defence against viruses

Bacteriophages have the potential to interfere with any industrial workflow that produces bacteria as an end product (such as starter cultures, live mucosal vaccines or probiotics) or use them as biocatalysts for the production of fermented food as well as of bioactive molecules (such as enzymes or metabolites). Viral infections can readily lead to cultures losses during industrial bioprocesses that rely on microbial metabolism. For this reason, companies (such as dairy and wine industries) are currently investing on research projects devoted to find new solutions in order to combat phage infections. In this regard, CRISPR systems might be exploited to design powerful methods to meet this need. For example, it would be possible to

boost the immunity of starter cultures by naturally or artificially adding new spacers into CRISPR loci [7, 8].

Specific silencing of endogenous genes

Functional genomics analyses of molecular components of CRISPR-Cas systems pointed out to a certain degree of analogy with those involved in the eukaryotic pathway of **RNA interference** (RNAi) [22]. Although this hypothesis needs to be experimentally confirmed, it is expected that enzymatic activities from CRISPR-Cas systems would represent a potential toolbox to **knock-down** specific gene functions. Interestingly, several independent studies on model systems, such as *Streptococcus pneumonia*, *Escherichia coli, Streptococcus pyogenes* and *Staphylococcus aureus*, have demonstrated that type II CRISPR systems can be reprogrammed to engineer bacterial genomes [23, 24]. For example, several *E. coli* genes encoding enzymes with β-lactamase activity were targeted, thus narrowing the spectrum of antibiotic resistance of the microbial population [23].

1.3 CRISPR-Cas systems of the themophilic archaeon *Sulfolobus solfataricus*

Unlike other *Archaea* and *Bacteria* showing one type of effector complex, genome sequence analysis revealed that ***S. solfataricus* P2** harbours: i) six CRISPR loci (named from A to F), showing two families of repeats; ii) four *cas6* genes, important for the maturation of the pre-crRNA; iii) three **acquisition cassettes** and iv) seven **interference cassettes** coding for different effector complexes, i.e. three of subtype I-A (**CASCADE complex**), one of subtype III-A (**CSM complex**) and three of subtype III-B (**CMR complex**) (Fig. 4). Altogether these features make this thermophilic archaeon an exceptional model for studying the interplay among different types of CRISPR-Cas systems [25]. Leader and repeat sequences are similar for the loci A and B (repeats family II), which have in their proximity two acquisition cassettes (Fig. 4). Similar gene cassettes are also located nearby the loci C, D and E (repeats family I), which share both repeat and leader sequences as well [27]. An exception is the locus F that lacks a leader sequence, thus being inactive in new spacer acquisitions (Fig. 4) [25, 26]. The association between a certain repeats family and an acquisition cassette led to speculate that specific Cas1/Cas2 couples are required in the adaptation phase of the related loci. Nevertheless, it has been demonstrated by *in vivo* experiments that no spacers were acquired by CRISPR loci A and B, although they have nearby a complete integration cassette (Fig. 4) [28]. However, their inactivity in new spacers integration is not correlated with an impaired capability of performing interference, indeed, loci A and B have been shown to be functional during the interference process [19, 26].

The biogenesis of mature crRNAs in *S. solfataricus* is potentially a complex process; indeed, four *cas6* paralogues have been identified in its genome (Fig. 4). For two of them (encoded by *sso1437* and *sso2004*), it has been shown that the cut of the crRNA occurs at 8-nucleotides from the 3'-end [25]. As for the *cas1*/*cas2* gene couples, the proximity of each *cas6* gene to a different *cas* operon encoding CASCADE and/or CSM complexes, reinforces the idea of a functional association between a distinct Cas6 and a certain ribonucleoprotein complex [26]. Nevertheless, another possibility is that single Cas6 proteins can process pre-crRNA derived from all different CRSIPR loci. This would explain why, in the study by Lintner and

colleagues, a single CASCADE complex was able to interact with crRNAs derived from all the CRISPR loci [29]. Mature crRNAs produced by Cas6 may be loaded into three types of effector complex, although a stable interaction between Cas6 and these complexes has never been demonstrated. Three interference cassettes on the *S. solfataricus* genome (Fig. 4, IA-1, - IA-2 and IA-3) encode for CASCADE complexes showing high similarity to those of the *E. coli* [29]. In all these operons, gene position is conserved, i.e. they harbour in the following order: i) a *csa5* gene, coding for the so-called 'small' subunit, followed by ii) *cas7* and *cas5* that probably assemble in a stable heterodimer and form the core of the complex and iii) *cas3'* and *cas3''* coding for the helicase and nuclease subunits of Cas3, respectively [29]. Moreover, the gene cluster lying between loci A and B contains a member of the *Cas8a2* family (Fig. 4, IA-1). As shown by many authors, the final target of these CASCADE complexes is DNA rather than RNA molecules [19, 28, 30].

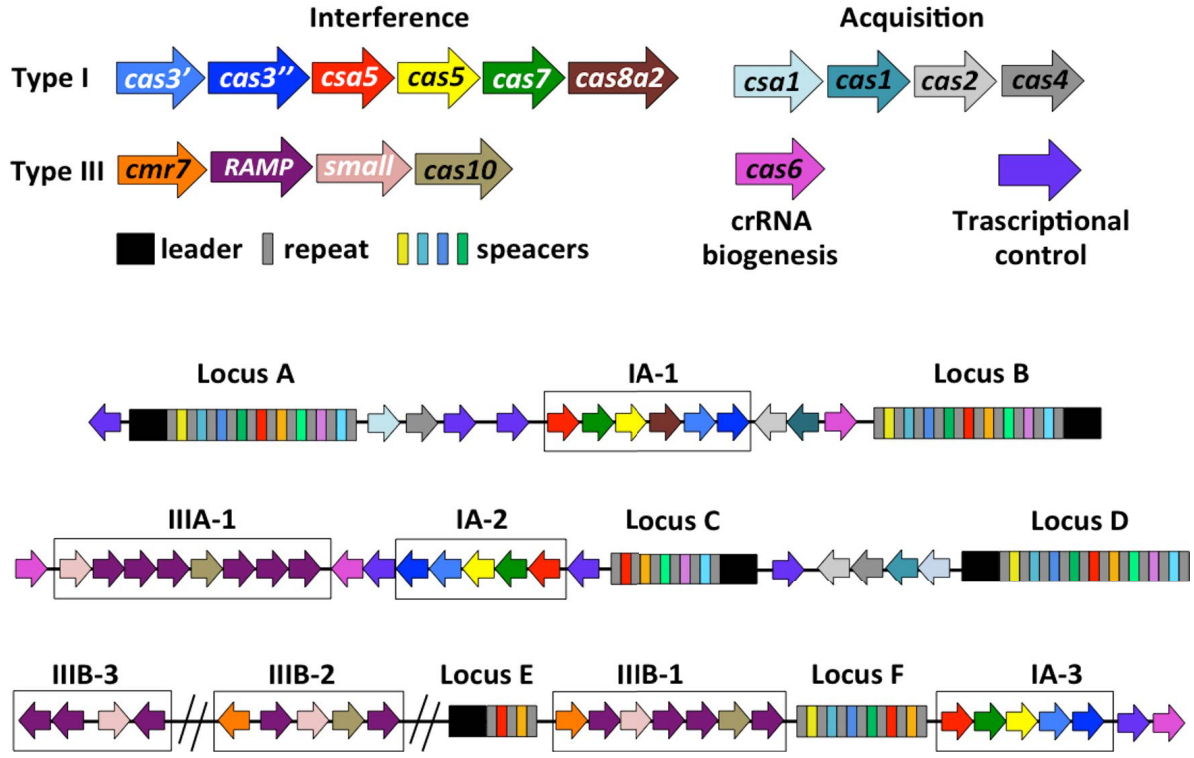


Figure 4. CRISPR-Cas systems organization in *S. solfataricus* P2. Genes encoding for effector complexes are boxed and color-coded accordingly to the above key.

Whereas in *E. coli* and other *Bacteria* the interference activity can be readily inactivated by the occurrence of one or very few mismatches between the spacer (on the crRNA) and the protospacer (on the foreign nuclei acid), for *S. solfataricus* an unexpected permissiveness has been experimentally demonstrated [30]. Indeed, its CASCADE complexes can tolerate up to 15 mismatched at the 3'-half of the crRNA (50-60 nt long), being still able to trigger 50% of target degradation. Nevertheless, likewise in *E. coli*, it was also confirmed the importance in the interference process of the **seed sequence** (i.e. the first 8-nucleotides at the 5'-end of the crRNA) [30]. *S. solfatarcus* encodes also for three CMR complexes, the structure of which consists of seven subunits Cmr1–Cmr7 (formerly known as RAMP proteins) that assemble around a central crRNA. Whereas Cmr7 is over-represented, the other subunits are likely present as single copies in the complex [25, 26]. In this subtype IIIB system, the

crRNA guides, in a sequence-specific manner, the CMR complex to cleave an RNA target.

The wide plethora of CRISPR loci, *cas* genes and interference complexes encoded by *S. solfataricus* suggests a specialization of these complexes in targeting heterogeneous genetic elements. Moreover, as mentioned above, pull-down assays have demonstrated that crRNAs coming from all different CRISPR loci can interact with a single type of CASCADE complex, thus forming “mixed” complexes [29]. This further expands the already intricate “arsenal” of defence strategies employed by *S. solfataricus*. Arguably, the extensive versatility of its CRISPR systems may be related to the huge abundance and diversity of viruses inhabiting the same niches.

In fact, archaeal viruses show morphtypes that have not been observed before for viruses infecting *Bacteria* and *Eukarya*. Furthermore, the lack of sequence identity between their genomes and those deposited in public database, have made even more difficult their classification. Indeed, eight novel viral families have been introduced in order to classify these viruses, i.e. *Fuselloviridae*, *Lipothrixviridae*, *Rudiviridae*, *Guttaviridae*, *Globuloviridae*, *Bicaudaviridae*, *Ampullaviridae* and *Clavaviridae* [ref 31]. However it is expected that there are still several unique archaeal viruses to be discovered. Among these, many analyses of environmental samples have shown that **spindle-shaped viruses** (SSVs) are abundant and occupy several niches, where they frequently outnumber head-tailed viruses. Notably, this unique virion morphotype seems to be a hallmark of viruses infecting *Archaea*, since it has never been observed for bacteriophages and eukaryal viruses [31]. To date, the *Fuselloviridae* family comprises nine members (SSV1, SSV2, SSV4, SSV5, SSV6, SSV7, SSV8, SSV9 and ASV1) isolated from several geographic locations. SSV1 is the most extensively characterized member of the family. Moreover, it is the only fusellovirus showing a UV-inducible life cycle, thus being an exceptional model study to investigate on virus-host interactions in *Archaea*. On the other hand, although SSV2 resemble SSV1 in virion morphology and genome organization, its replication is triggered by the physiological state of the host cell rather than by an external stimulus as for SSV1. A complete dissertation on fuselloviruses is available in the review appended below.

1.4 Molecular biology of fuselloviruses and their satellites

Extremophiles (2014) 18:473–489
DOI 10.1007/s00792-014-0634-0

REVIEW

Molecular biology of fuselloviruses and their satellites

Patrizia Contursi · Salvatore Fusco · Raffaele Cannio · Qunxin She

Received: 18 September 2013 / Accepted: 31 January 2014 / Published online: 23 February 2014

Abstract Fuselloviruses, also known as *Sulfolobus* Spindle-shaped viruses (SSVs), are "lemon"- or "spindle"-shaped double-stranded DNA viruses. Among them, SSV1, SSV2 and the satellite viruses pSSVx and pSSVi have been investigated at the structural, genetic, transcriptomic, proteomic and biochemical levels, thus becoming models for dissecting DNA replication/gene expression in *Archaea*. Important progress has been made including elucidation of temporal genome expression during virus infection and induction of replication, SSV1 lysogeny maintenance as well as differentially expression of pSSVx replicase. Future researches focusing on these model systems would yield insightful knowledge of life cycle and DNA replication of fuselloviruses.

Keywords Fuselloviridae · *Sulfolobus* · Satellite viruses · Transcription regulation

Communicated by S. Albers.

P. Contursi (✉) · S. Fusco
Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, Naples, Italy
e-mail: contursi@unina.it

R. Cannio
Unity of Rome, c/o Department of Chemistry, Institute of Biomolecular Chemistry (ICB), Sapienza University of Rome, Rome, Italy

Q. She (✉)
Department of Biology, Danish Archaea Centre, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark
e-mail: qunxin@bio.ku.dk

Introduction

Prokaryotes are the most abundant organisms on Earth but they are outnumbered tenfold by the viruses infecting them. Thus, prokaryotic viruses are the most predominant biological entities on Earth (Bamford 2003; Bamford et al. 2005; Le Romancer et al. 2007; Krupovic et al. 2011; Forterre and Prangishvili 2013). Shortly after the recognition of the *Archaea* (formerly *Archaebacteria*) (Woese et al. 1978; Albers et al. 2013), research focused on isolation of viruses and virus-like particles from extremely hot, low pH or hypersaline niches (Guixa-Boixareu et al. 1996; Oren et al. 1997; Rice et al. 2001; Rachel et al. 2002; Häring et al. 2005; Porter et al. 2007; Bize et al. 2008; Redder et al. 2009; Sime-Ngando et al. 2011; Snyder and Young 2011; Pietilä et al. 2013; Porter et al. 2013). Intriguingly, archaeal viruses exhibited completely novel and unique morphologies (Prangishvili and Garrett 2005; Lawrence et al. 2009; Pina et al. 2011; Ackermann and Prangishvili 2012; Peng et al. 2012; Prangishvili 2013) and appear to be adapted to the environments from which their hosts thrive (Snyder et al. 2003; Ortmann et al. 2006; Lawrence et al. 2009; Atanasova et al. 2012; Prangishvili 2013; Bartolucci et al. 2013), thus pointing to a process of host–virus coevolution within each biogeographic context (Held and Whitaker 2009). For example, for the viruses that infect hyperthermophilic organisms, their virions are also extremely thermostable and viral infection occurs most effectively at the optimal growth temperatures of their hosts (Schleper et al. 1992; Zillig et al. 1996; Ceballos et al. 2012; Quemin et al. 2013; Fu and Johnson 2012). This has led to the hypothesis that archaeal viruses are very ancient and may have preceded the separation of the three Domains of life (Prangishvili 2003; Forterre 2006; Ortmann et al. 2006) such that each Domain had already been

accommodated with a fraction of the existing diverse virus population when it was formed (Prangishvili et al. 2006b; Koonin et al. 2006).

Compared to bacteriophages that mainly show the typical head-to-tail morphotypes, archaeal viruses show a remarkable diversity in morphology, which imposed the establishment of a number of new virus families (Pina et al. 2011; Peng et al. 2012; Prangishvili 2013). To date, all known archaeal viruses contain double-stranded DNA (dsDNA) as genetic material except for two recent reports of single-stranded DNA viruses isolated from a haloarchaeaon (Pietilä et al. 2009) and from a crenarchaeon (Mochizuki et al. 2012). It is also worth mentioning that putative archaeal RNA viruses have been detected in a metagenomic study but such biological entities still remain to be obtained from an archaeal hosts (Bolduc et al. 2012; Stedman et al. 2013).

The majority of known archaeal viruses infect organisms belonging to *Creanarchaea*, including, but not limited to, members of the genera of *Sulfolobus*, *Acidianus*, *Thermoproteus*, *Aeropyrum* and *Pyrobaculum* (Prangishvili et al. 2006a, Prangishvili 2013). This attributes at least partly to the major effects of isolating archaeal viruses from thermophilic environments in which these archaeal organisms flourish. Archaeal virus research was initiated by Wolfram Zillig, Wolf-Dieter Reiter and colleagues (Reiter et al. 1987a, b, 1988a, b, 1989; Zillig et al. 1994, 1996; Albers et al. 2013) and followed by David Prangishvili, Roger Garrett, Mark Young and colleagues (Prangishvili et al. 2001; Prangishvili and Garrett 2004, 2005; Wiedenheft et al. 2004; Pina et al. 2011). In parallel, there were focused researches on developing *Sulfolobus* model organisms for studying novel biological principles in *Archaea*. In particular genome sequencing of *Sulfolobus solfataricus* P2 (She et al. 2001b) and *Sulfolobus islandicus* REY15A, HVE10/4 and LAL14/1 (Guo et al. 2011; Jaubert et al. 2013) as well as the development of microarrays containing probes of all host genes and of several viral genomes has allowed archaeal host–virus interaction to be investigated at genome scale using DNA microarrays (Fröls et al. 2007; Ortmann et al. 2008; Okutan et al. 2013; Ren et al. 2013) and RNA sequencing (Quax et al. 2013). These developments render *Sulfolobus* host–virus systems good models for studying molecular biology of archaeal viruses. Nevertheless, gene and protein sequences from crenarchaeal viruses are often orphans, i.e., they do not show detectable sequence similarity to proteins present in the public databases. As a result, functional annotation has not been performed for these gene products so far (Prangishvili et al. 2006a, b; Menon et al. 2010). For this reason, the comprehension of the fundamental viral processes in archaea, such as uptake to host cells, transcriptional regulation, genome replication virus assembly and release, is still at a stage of infancy.

Nearly all crenarchaeal viruses fall into eight families including *Fuselloviridae*, *Lipothrixviridae*, *Rudiviridae*, *Guttaviridae*, *Globuloviridae*, *Bicaudaviridae*, *Ampullaviridae* and *Clavaviridae* (Prangishvili et al. 2006a; Mochizuki et al. 2010; Pina et al. 2011; Peng et al. 2012) and two other families whose approval is still pending at the International Committee on Taxonomy of Viruses (ICTV) (Prangishvili 2013). Research on *Fuselloviridae* has been centered on a few model viruses. SSV1 (*Sulfolobus* spindle-shaped virus 1) is the first fusellovirus to be characterized (Martin et al. 1984), representing the prototype of this family for physiological, genetic and transcriptional studies (Reiter et al. 1987a, b; Reiter et al. 1988a, b, 1989; Schleper et al. 1992; Stedman et al. 1999; Fröls et al. 2007; Iverson and Stedman 2012). The second studied fusellovirus is SSV2, which shows distinct life cycles in the natural versus foreign hosts (Stedman et al. 2003; Contursi et al. 2006). Also investigated are the interactions of SSV2 with virus satellites, which are plasmid–virus hybrids that form virions by hijacking the viral packaging machinery of a helper virus (Arnold et al. 1999; Wang et al. 2007; Ren et al. 2013). Although SSV8 (aka SSV RH) has not been studied physiologically, a few proteins encoded in this viral genome have been characterized to reveal functional diversity of viral proteins. Interestingly, integration of SSV9 (aka SSV K1) occurs not only at tRNA locations, thus representing an exception to the evolutionary conserved process of integrating viruses and phages (Wiedenheft et al. 2004; Prangishvili 2013).

In this review, we focus on the current knowledge of the molecular and physiological features of the *Fuselloviridae* and how the acquired knowledge significantly expanded our understanding of archaeal biology.

An overview of fuselloviral morphology, diversity and evolution

To date, *Fuselloviridae* comprise nine members derived from four different geographic locations. The founding member SSV1 was isolated from a hot spring in Beppu, Japan (Martin et al. 1984). Subsequently, five fuselloviruses (SSV2, SSV4, SSV5, SSV6, and SSV7) were obtained from different geothermal environments in Iceland (Stedman et al. 2003; Redder et al. 2009). Moreover, Mark Young and colleagues have isolated several new archaeal viruses including SSV9 from a hot spring sample from Kamchatka (Russia) and SSV8 from the Yellowstone National Park (USA; Wiedenheft et al. 2004). The only non-*Sulfolobus* fusellovirus isolated thus far is ASV1 (*Acidianus* spindle virus 1), which was obtained from

Extremophiles (2014) 18:473–489

Table 1 Features of all known fuselloviruses

Virus name	Sampling site	Genome size (bp)	ORFs number	NCBI number	Reference
SSV1	Japan	15,465	35	NC_001338	Palm et al. 1991
SSV2	Iceland	14,796	35	NC_005265	Stedman et al. 2003
SSV4	Iceland	15,135	34	EU030938	Redder et al. 2009
SSV5	Iceland	15,330	34	EU030939	Redder et al. 2009
SSV6	Iceland	15,684	33	NC_013587	Redder et al. 2009
SSV7	Iceland	17,602	33	NC_013588	Redder et al. 2009
SSV8	USA	16,473	37	NC_005360	Wiedenheft et al. 2004
SSV9	Kamchatka	17,385	31	NC_005361	Wiedenheft et al. 2004
ASV1	USA	24,186	38	NC_013585	Redder et al. 2009

Acidianus brierleyi, a thermoacidophilic archaeon isolated from a hot spring in the Yellowstone National Park, USA (Redder et al. 2009) (Table 1).

Whereas all the known *S. solfataricus* strains and some icelandic species such as *S. islandicus* HVE10/4 can propagate efficiently most of the SSV viruses, SSVs infection is not effective on *S. islandicus* RENH1, *S. acidocaldarius* and *S. tokodaii* species. On the other hand, SSVs display significant variations in their host ranges with SSV1 exhibiting the narrowest and SSV8 exhibiting the broadest host range, respectively. Therefore, SSV infectivity and *Sulfolobus* susceptibility are not related to the geographical context from which the hosts and viruses were isolated (Held and Whitaker 2009; Ceballos et al. 2012).

These nine genomes provide geographically distinct isolates that have been used in comparative genomic and morphological analyses (Wiedenheft et al. 2004; Redder et al. 2009).

The genome of *Fuselloviradae* contains dsDNA ranging from ca. 15–22 kb in size that is encased by 2–3 structural proteins (VP1, VP2 and VP3) to form spindle-shaped viral particles (Fig. 1a). In particular, for SSV1 it has been experimentally demonstrated that the capsid is formed by the interaction of VP1, VP2 and VP3 proteins (Reiter et al. 1987a). Nevertheless, most of fuselloviral genomes contain only genes encoding for VP1 and VP3 (Redder et al. 2009). Under electron microscope, virions of most known fuselloviruses appear as 55−60 × 80−100 lemon-shaped particles (Stedman et al. 2003; Wiedenheft et al. 2004; Redder et al. 2009). However, two fuselloviruses SSV6 and ASV1 show pleomorphic and malleable morphologies, varying from a pear-like shape to an elongated cigar shape (Fig. 1b, c). A close examination of these virions revealed filaments attached to one of the pointed ends (Martin et al. 1984; Palm et al. 1991; Schleper et al. 1992; Wiedenheft et al. 2004; Redder et al. 2009), which are likely to be implicated in anchoring the viruses to the host membrane upon infection as well as to cause virus clustering into rosette formations as seen in cultures of almost all the isolates.

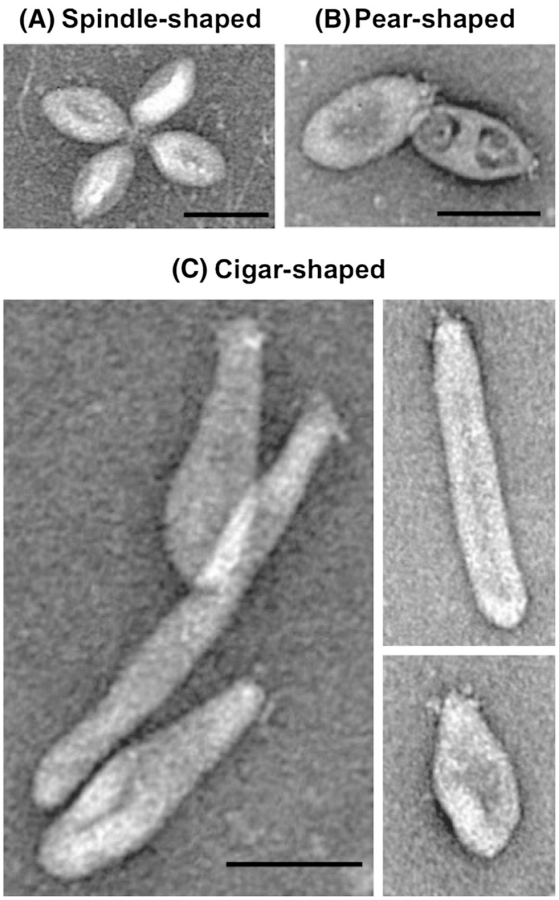


Fig. 1 Electron microscopy images of *Fuselloviridae* virions. **a** Exemplification of the typical lemon-/spindle-shaped morphology, which is displayed by SSV1, SSV2, SSV4, SSV5, SSV7, SSV8 and SSV9. **b**, **c** Viral particles of ASV1 and SSV6 exhibiting a wide range of morphotypes varying from the pear shape to the elongated cigar shape. *Scale bars* 100 nm. Reproduced with permission from Redder et al. 2009

Two different variants in the number and in the structure of these filaments have been discovered: multiple, thin, quite sticky filaments, readily linking virions are present in all

the *Fuselloviridae* with the exception of SSV6 and ASV1. The latter two, indeed, carry 3–4 thick and slightly curved short filaments, forming a crown around the virion tip. It seems that the difference in the numbers and structure of the terminal ends in the two morphotypes correlates with the presence of two different variants of the longest ORF (open reading frame) present on the fuselloviral genomes, i.e., the SSV1-C792 or the SSV6-B1232 module, respectively (Redder et al. 2009). Despite the pleomorphism in virions, all fuselloviral genomes show a high similarity in gene synteny indicating a clear relationship among the viral isolates. As is typical of many viral genomes, the predicted ORFs are tightly arranged on the genome with very small intervening non-coding sequences. Sequence alignments of most of the ORFs show little or no sequence similarity in the public databases. However, most of the fuselloviral ORFs are homologoues to each other and are also collinearly organized (Wiedenheft et al. 2004). A very illustrative picture of the fuselloviral genomes comparison is reported in the paper of Redder et al. 2009. Thirteen core genes are conserved among all members, which are also considered as the minimal genetic signature of this virus family and may represent viral functions common to all fuselloviruses, despite their different geographic isolation (Wiedenheft et al. 2004; Redder et al. 2009; Held and Whitaker 2009; Krupovic et al. 2012). The core set of ORFs includes *vp1* and *vp3* coding for the capsid proteins, three putative transcriptional regulators, six orphan proteins and a type I tyrosine recombinase that facilitates provirus formation (Argos et al. 1986; Serre et al. 2002). This common set of 13 ORFs shared by all the nine isolates is not contiguous. Nevertheless, it is noteworthy that conserved ORFs are mainly confined to one half of the SSV virus genomes and are arranged in the same orientation. Conversely, the other half of the genomes is more divergent and ORFs unique to any one fuselloviral isolates may reflect their evolutionary history, geographic isolation, requirements for replication in their specific hosts or adaptations to features of their thermal environments (Wiedenheft et al. 2004; Redder et al. 2009).

In the less conserved half of the genomes, a clustering of ORFs is evident containing cysteine codons that are unusual for hyperthermophiles (Stedman et al. 2003). Further analyses have indicated that there is a general abundance of disulfide bonds in the intracellular proteins of hyperthermophilic viruses but a decreased cysteine content in their membrane proteins, which are often late gene products of hyperthermophilic viruses (Menon et al. 2008). Therefore, it has been reasoned that such a genomic organization for fuselloviruses more likely reflects the clustering of late genes of these viruses rather than a fusion event of two genomes with distinct histories (Menon et al. 2008).

Apparently there are evolutionary constraints to maintain gene synteny for fuselloviruses, which can be exemplified with the comparison between SSV1 and SSV2. Although isolated from hot springs of geographically distant locations, these two viruses show collinear genomic organization, with the majority of their ORFs being homologous to each other (Stedman et al. 2003). Nevertheless, SSV1 and SSV2 exhibit interesting differences in physiology, life cycle and relationship with the relative hosts (see below), which reflect evolutionary adaptation to their hosts and/or to the environment.

A mechanism has been proposed for the evolution of fuselloviruses invoking homologous recombination between two integrated viruses. The hypothesis is based on the fact that closely related fuselloviruses possess multiple highly similar or even identical regions. For instance, SSV4 and SSV5 share an almost identical 7.9 kb DNA region. If co-integration of two viral genomes into the same site generates concatemers of different viruses, their excision can occur at multiple positions leading to diverse virus variants (Redder et al. 2009).

SSV1: the prototype of the *Fuselloviridae*

SSV1–host interaction

SSV1 was first isolated as a plasmid present in *S. shibatae* B12 (Yeats et al. 1982). Then virus-like particles (VLP) were observed and their production was found to be UV-inducible (Martin et al. 1984). Subsequently, it was found that the VLP infected *S. solfataricus* (Schleper et al. 1992), thereby demonstrating the viral nature for SSV1.

SSV1 episomal DNA is stably maintained in host cells in three different forms: the DNA is either positively or negatively supercoiled or relaxed double stranded (Nadal et al. 1986). Whereas the integrated form might be used as a template for transcription, the positively supercoiled is likely encased in the viral particle, while the negatively supercoiled might be used for genome replication (Snyder et al. 2003). SSV1 is the only known member of the *Fuselloviridae* family that shows an UV-inducible gene expression and genome replication. Cells are not lysed upon virus particle release and recover their growth rate as well as their lysogenic state within few hours (Schleper et al. 1992).

The complete genome of SSV1 was sequenced (Palm et al. 1991) and the integrase that furthers site-specific integration has been biochemically characterized (Muskhelishvili et al. 1993; Serre et al. 2002; Zhan et al. 2012). The integration occurred site specifically at an arginyl-tRNA gene and an intact host gene is maintained after integration (Reiter et al. 1989; Muskhelishvili et al. 1993;

Extremophiles (2014) 18:473–489

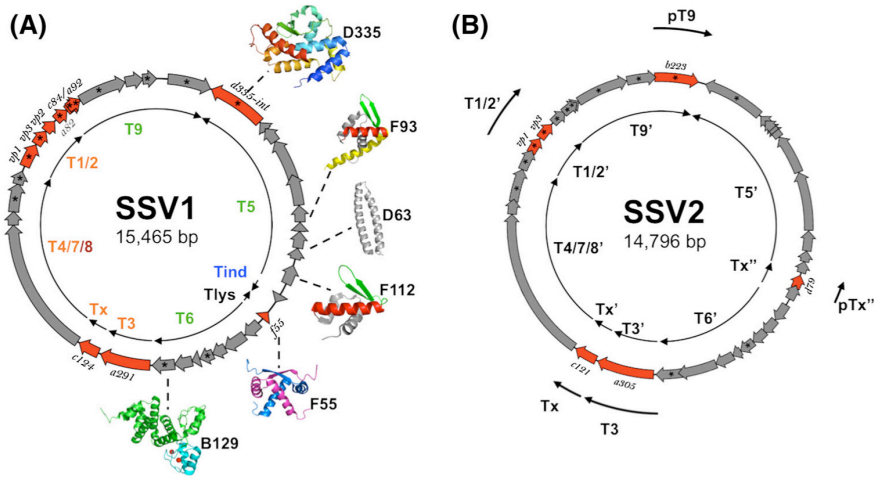


Fig. 2 Genomic map of SSV1 and SSV2. ORFs labeled with the *asterisk* belong to the core set of 13 genes conserved in all *Fuselloviridae* (Redder et al. 2009) ORFs in *red* are those expressed in the lysogenic state of both SSV1 and SSV2. *Inner black arrows* represent the identified transcripts of the fuselloviruses. The transcriptional map has been experimentally determined for SSV1 (Reiter et al. 1987b; Fröls et al. 2007) and deduced from microarray data for SSV2 (Ren et al. 2013). **a** The three-dimensional structures of SS proteins (determined for D63, F93, F112, B129 and D355 C-terminal domain and predicted for F55) are shown. **b** T transcriptional map of SSV2 in stable *S. solfataricus* transfecta deduced from the expression pattern of the ORFs is indicated by *outer black arrows*. pT_9 and pT_x'' stand for partial T_9 and transcripts

Serre et al. 2002). Conversely, since the attachment site is located within the *int* gene, integration results in the partitioning of the gene that inactivates the expression of the encoded integrase. This mechanism accounts for generating gene capture in archaeal genomes (She et al. 2001a). Interestingly, the integrase gene has been shown to be dispensable for virus replication as well as for its spreading into *Sulfolobus* cultures, thus demonstrating that the viral integration is an optional step in the replication of SSV1 and this is probably true for all fuselloviruses (Clore and Stedman 2007). Although Clore and Stedman showed that the wild-type SSV1 outcompeted the virus without an integrase gene, a positive pressure for the maintenance of these genes has been hypothesized, since integrase-encoding sequences are widespread in fuselloviruses genomes as well as in other integrative genetic elements (She et al. 2001a, 2002, 2004; Peng et al. 2000; Peng 2008; Cortez et al. 2009). Obviously, integrase genes have played an important role in horizontal gene transfer and genome evolution.

SSV1 gene expression: insights into promoters and terminators of archaeal transcription

The regulation of gene expression on SSV1 was studied immediately after the isolation of this genetic element. First, constitutive and UV-inducible transcripts were identified, including an UV-inducible transcript (T_{ind}), and nine constitutive messengers (T_1–T_9) (Reiter et al. 1987a, 1988a, Zillig et al. 1988). More detailed analysis usi microarrays showed that the expression of these vi transcripts is temporally regulated such that they fall i four distinct classes: (1) UV-inducible (T_{ind}), (2) early (T_6 and T_9), (3) late (T_3, $T_{4/7}$), and (4) late-extended ($T_{4/}$ messengers. Moreover, an additional late monocistro transcript was detected and mapped (T_x) (Fig. 2a). T fashion of regulation is reminiscent of that adopted many bacteriophages and eukaryotic viruses (Fröls et 2007).

Then, transcriptional start sites (TSSs) were identif for all the SSV1 transcripts (Reiter et al. 1988a), wh facilitated the identification of promoter elements of SS Two conserved sequence motifs, including a TATA-bo like hexanucleotide sequence (box A, TTTAAA) that centered about 26 nucleotides upstream of the TSS an trinucleotide sequence (box B; consensus sequence TG that is localized nearby the TSS, were identified (Rei et al. 1988a). Strikingly, these elements resemble those the eukaryotic basal gene promoters recognized by eukaryotic RNA polymerase II (Reiter et al. 1988a) a this is consistent with the results of the similar subu composition of the eukaryotic and archaeal RNA po merases unraveled from another study (Huet et al. 198

Analyzing transcription termination signals of th SSV1 transcripts has identified a conserved motif with consensus TTTTTYT (Reiter et al. 1988b). Once ag these sequences resemble the pyrimidine-rich elements t are implicated in transcription termination by eukaryo

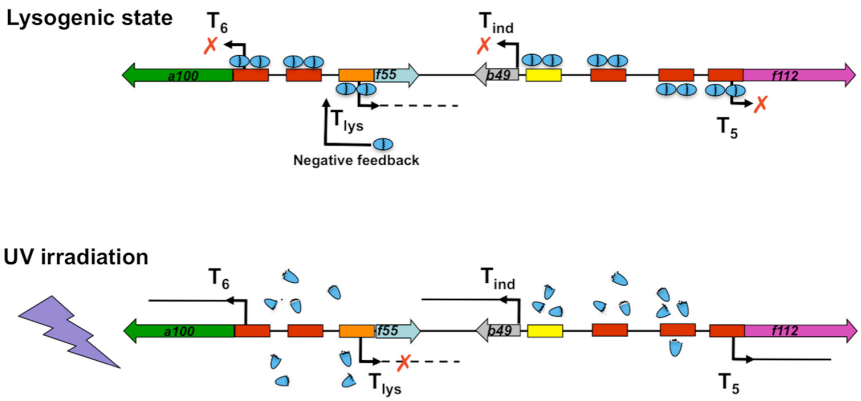


Fig. 3 Model of the F55 interaction at its binding sites as modified from Fusco et al. 2013. In the lysogenic state, F55 (*cyan ovals*) binds as dimers to the target sequences in the promoters of T_5, T_6, T_{ind} (*red and yellow boxes*) as well as to its own promoter (*orange box*). As shown by *red crosses*, transcription of T_5, T_6 and T_{ind} is locked, while the expression of its own gene is progressively turned off following a negative feedback control. Upon UV irradiation, F55 might be degraded and/or inactivated and consequently transcription of the early T_{ind}, T_5 and T_6 transcripts is unlocked

RNA polymerase I (Grummt et al. 1985; Reiter et al. 1988b).

More recently, our research has gained some insights into the molecular components and the mechanisms underpinning the maintenance of the SSV1 lysogeny. Transcription analysis of an SSV1-harboring *S. solfataricus* strain, in the absence of the UV stimulus (lysogenic state), showed the expression of a novel SSV1 transcript, named T_{lys}. This messenger is expressed from a genomic region located nearby the UV-inducible T_{ind} (Fig. 2a), which is in turn totally repressed under the same conditions. T_{lys} encodes for a 6.3-kDa protein, F55, which is able to interact specifically at operator sequences in the promoters of the early transcripts (T_{ind}, T_5 and T_6) as well as of its own messenger. F55 is predicted to bear the ribbon–helix–helix fold typical of negative transcription regulators. Therefore, by binding to its target sequences, F55 might exert down-regulation of the early genes in the lysogenic state. By analogy with the cI repressor of lambda phage, F55 is possibly degraded and/or inactivated upon UV irradiation, thus unlocking the transcriptional circuit of the early genes. The hypothesis of SSV1 induction and the molecular components involved in the SSV1 life cycle are illustrated in Fig. 3 (Fusco et al. 2013).

Structural and functional analyses of SSV1 proteins

There are 34 predicted ORFs on the SSV1 genome, encoding protein products that range from 6 to 86 KDa and about 75 % of these ORFs have not been reliably identified by bioinformatic approaches. Function is only known for the viral structural proteins (VP1, VP2 and VP3) (Reiter et al. 1987a) and for the integrase D335, whose structural analysis of the C-terminal catalytic domain has revealed that it possesses a core fold similar to those of type I tyrosine recombinases of both bacterial and eukaryal origin. In vitro studies showed that this enzyme is capable of transferring a phosphodiester bond from host to viral attachment sites (attA and attP, respectively) as well as of carrying out the reverse reaction (Muskhelishvili et al. 1993; Serre et al. 2002; Eilers et al. 2012; Zhan et al. 2012). VP2 is thought to function as a small, packaged, DNA-binding protein (Stedman et al. 2003; Wiedenheft et al. 2004). Two additional components that copurify with the viral particles have also been identified, C792 and D244 (Menon et al. 2008). C792 is a predicted membrane protein serving probably a structural role by generating filaments at the tail end of virions implicated in host receptor (Redder et al. 2009), while D244 is a soluble protein of unknown function (Menon et al. 2010) whose homolog from the SSV8 virus has been studied at structural level revealing a possible role in DNA replication, repair, or recombination (see below).

B251 exhibits limited similarity to DnaA (Koonin 1992). Furthermore, molecular modeling allowed structure prediction for four SSV1 proteins. While E51, C80 and F55 might adopt the ribbon–helix–helix fold, A45 and A79 carry a C2H2 zinc finger-like motifs (Prangishvili et al. 2006b). However, with the exception of F55, their DNA-binding activity remains to be tested and therefore their role in the virus life cycle is murky.

The general lack of insights into protein functions from sequence analyses stimulated the X-ray crystallographic approach to determine structures of SSV1 proteins. Indeed, structural analysis of viral proteins provides an important alternative approach to obtain a deeper comprehension of

the life cycle of SSV1 and crenarchaeal viruses in general (Lawrence et al. 2009; Krupovic et al. 2012). Beside the C-terminal domain of the integrase, 3D structures are available for 4 SSV1 ORFs, all encoding putative transcriptional factors. In the structure of D63 (Kraft et al. 2004a), the helix-turn-helix fold resembles that of the "repressor of primer" (ROP), an adaptor protein that regulates colE1 plasmid copy number in *E. coli* (Helmer-Citterich et al. 1988). F93 shows a fold similar to proteins belonging to the MarR (Alekshun et al. 2001; Di Fiore et al. 2009; Fiorentino et al. 2011) and SlyA (Wu et al. 2003) subfamilies of winged-helix transcription regulators (Kraft et al. 2004b). The winged-helix protein F112 shows structural homologies with DP2 (Gibrat et al. 1996), a regulator that plays a central role in the eukaryotic cell cycle (Menon et al. 2008). Finally, B129 encodes a C2H2 Zinc finger transcription factor (Lawrence et al. 2009).

However, functions of these SSV1 regulators as well as of other gene products of unknown function remain to be dissected by genetic analysis. Stedman and colleagues have made first attempts to reveal essential genes of the SSV1 virus (Stedman et al. 1999).

In a subsequent study, three ORFs were deleted from the SSV1 genome, i.e., the universally conserved *b129*, the well-conserved *d244* and the poorly conserved *vp2*. The inactivation of the ORF encoding for the predicted transcriptional regulator B129 resulted in a loss of infectivity, while deletion of the one encoding the predicted DNA-binding protein VP2 yielded viable viruses that were indistinguishable from the wild-type. Interestingly, a new phenotype was observed for a *d244*-defective mutant, i.e., it was able to slow down the growth of the host in liquid culture (Iverson and Stedman 2012).

SSV2 and its satellites

Life cycle

SSV2 and its hosts have been subjects of physiological characterization for the study of fusellovirus–hosts interactions. SSV2 was isolated from *S. islandicus* REY 15/4 together with pSSVx, a virus satellite (Arnold et al. 1999) and as for SSV1, this virus also replicates in *S. solfataricus* P2 and infected cells carry a provirus in the chromosomes site specifically integrated at a $tRNA^{Gly}$ gene (Contursi et al. 2006).

Differently from SSV1, which exhibits a dramatic UV-inducible virus production, SSV2 apparently does not and shows only a modest induction of 2-fold (Stedman et al. 2003). Instead, SSV2 exhibits a physiological induction of virus replication dependent on the growth (Contursi et al. 2006). Thus, both viruses are interesting models for studying molecular biology of archaeal viruses.

SSV2 is very similar to SSV1 in virion morphology and in gene synteny and in sequence, among a total of 34 ORFs of SSV2, 26 show significant similarity to SSV1 ORFs (Stedman et al. 2003). SSV2 carries a direct repeat of 62 base pairs, with the core sequence repeated 4.5 times, which could be regarded as a hallmark of the SSV2 genome as it is absent from all other known fuselloviruses (Stedman et al. 2003, Wiedenheft et al. 2004). In comparison, the T_{ind} and T_{lys} transcripts are the hallmarks of SSV1 (Fig. 2a). Together, these differences may account for the lack of ultraviolet induction of SSV2 and for an elevated level of virion production detected in the lysogenic state of SSV2-infected *S. solfataricus* cells (see below).

Another striking difference is that SSV2 shows a broader host range compared to SSV1. Among the Icelandic *Sulfolobus* species susceptible to SSV2 infection (Ceballos et al. 2012) there is *S. islandicus* REY15A (Contursi et al. 2006), a strain isolated from the same enrichment culture of its natural host (Arnold et al. 1999). *S. islandicus* 15A and *S. solfataricus* P2 have been chosen to study the virus–host interaction in closely related and foreign systems, respectively.

The life cycle of SSV2 in the natural host is characterized by induction of virus replication at a later growth stage. It has been hypothesized that the induction is triggered by an unidentified signal molecule or by a transcriptional factor that exhibits growth phase regulation. Viral DNA replication upon induction shows the following features: (1) a steep increase of the SSV2 copy number (about 50 folds) within 4 h; (2) a concurrent inhibition of the host growth at a late exponential growth phase; (3) a consistent packaging and extrusion of the viral particles and (4) the reversibility of the replication induction and of the growth inhibition (Contursi et al. 2006).

The induction of SSV2 replication also occurs for the infected *S. islandicus* REY15A. However, the viral replication is not induced in SSV2-infected *S. solfataricus* cells, suggesting that the SSV2 physiological induction has resulted from mutual interactions that only exist for SSV2 and certain *S. islandicus* strains.

The interplay between SSV2 and *S. solfataricus* during the process of the infection has been studied through global analysis of the gene expression of the viral and host genomes in primary infected cells (Ren et al. 2013). In this case, the activation of SSV2 genes follows a chronological scheme based on a distributive pattern with all the genes transcribed within 7.5 h post-infection (Fig. 2b). More investigations are required to illustrate the involved mechanisms.

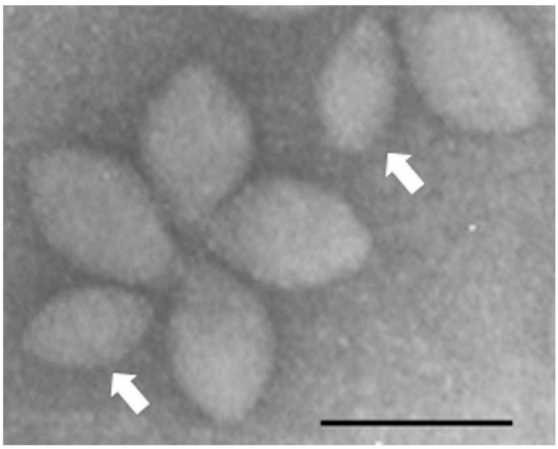


Fig. 4 Electron microscopy image of SSV2 and pSSVi viral particles. *White arrows* point to the smaller pSSVi-containing virions. *Scale bar* 100 nm. Reproduced with permission from Wang et al. 2007

Two hybrid virus/plasmid genetic elements, denoted as pSSVx (Arnold et al. 1999) and pSSVi (Wang et al. 2007), are packaged into small spindle-shaped virion particles (60 nm x 40 nm) in the presence of SSV1 or SSV2 as helpers (Fig. 4), thus representing the only two satellite-helper virus systems known in the archaeal domain. Interesting insights have been gained into interactions between these genetic elements as well as between them and their hosts.

Interplays between SSV2 and virus satellites

Archaeal virus satellites were first identified by analyzing virus-like particles present in an *S. islandicus* REY15/4 culture in which spindle virions of two distinct sizes were observed: the larger and the smaller ones contained the genome of SSV2 and pSSVx, respectively. In this system, SSV2 is an ordinary virus and acts as a helper to pSSVx such that the latter hijacks the virus packaging machinery for its own virion formation (Arnold et al. 1999; Stedman et al. 2003).

pSSVx has seven ORFs, 3 of which are viral-like while the remaining 4 are plasmid-like. The viral-like genes *orf154* and *orf288* exhibit high sequence similarity to two of the 13 core genes well conserved among all known SSV viruses (Redder et al. 2009) and the third gene, *orf-c68*, is only present in virus satellites and integrated elements (Contursi et al. 2007; Wang et al. 2007). Strikingly, pSSVx devotes ca. half of its coding capacity to code for *orf892*, a homolog of the putative replicase of pRN1 and for three additional plasmidic ORFs, ORF51 (a putative copy number control protein), ORF91 (containing a putative zinc-binding motif) and ORF76 (a putative leucine-zipper protein), the latter three belonging to the category of DNA-binding proteins/transcriptional regulators (Fig. 5a). Taken together, this suggests that pSSVx replicates its genome in a fashion similar to plasmids of the pRN family (Keeling et al. 1996, 1998; Peng et al. 2000; Kletzin et al. 1999).

Putative replicases encoded in the pRN family have only been characterized for *S. islandicus* pRN1 plasmid (ORF904). The protein is a multiple domain enzyme: the N-terminal domain shows primase and DNA polymerase activities and the C terminus contains the winged-helix DNA-binding domain displaying helicase/ATPase activity (Lipps et al. 2003). The latter is necessary for DNA unwinding at the replication origin. To date, a detailed mechanism remains to be illustrated for the DNA replication of pSSVx and other plasmids of this family.

The second virus satellite, the pSSVi (Fig. 5b), has been identified in the genome of a laboratory strain of *S. solfataricus* P2 where it was integrated into the $tRNA^{Arg}$ gene. Integration is catalyzed by an integrase encoded by its own genome and the integrated form is stably maintained and propagated over several generations (Wang et al. 2007). This condition is perturbed by the co-presence of SSV2 (or SSV1), leading to the active replication of the pSSVi episomal form and to its packaging into spindle-shaped virus particles (Ren et al. 2013). Unlike pSSVx, its putative replicase gene is clearly different from those encoded by pRN plasmids (Wang et al. 2007). Characterization of the encoded enzyme shows that it is a DNA helicase belonging to the superfamily 3 helicase which interacts with the host primase in vitro, suggesting it could function as the replication initiator of pSSVi (Guo and Huang 2010). Interestingly, among the integrated elements identified in archaeal genomes, pSA2 and pST1 exhibit a genetic organization similar to that of pSSVi, suggesting they could also represent integrated virus satellites in *Sulfolobus* genomes (She et al. 2002, 2004; Cortez et al. 2009).

To date, mechanisms of virions satellite formation in *Archaea* are mainly inferred from bioinformatic analyses of their genomes. Initially, three viral-like ORFs of pSSVx, ORF-c68, ORF154 and ORF288, were supposed to be implicated in the virion formation. However, the absence of any homologues of the pSSVx ORF154 and ORF288 in the pSSVi genome suggests that these two ORFs are not essential for the formation of satellite virions (Wang et al. 2007). This raises the possibility that the packaging machinery of a fusellovirus recognizes a sequence motif to start virus packaging. Interestingly, the sequence motif 5′-AAGGGAAANAGNA-3′ is present in the genomes of pSSVx, pSSVi and SSV2 (989–1001, 2412–2424, 1406–1418 bp on their linear map, respectively), which is

Extremophiles (2014) 18:473–489

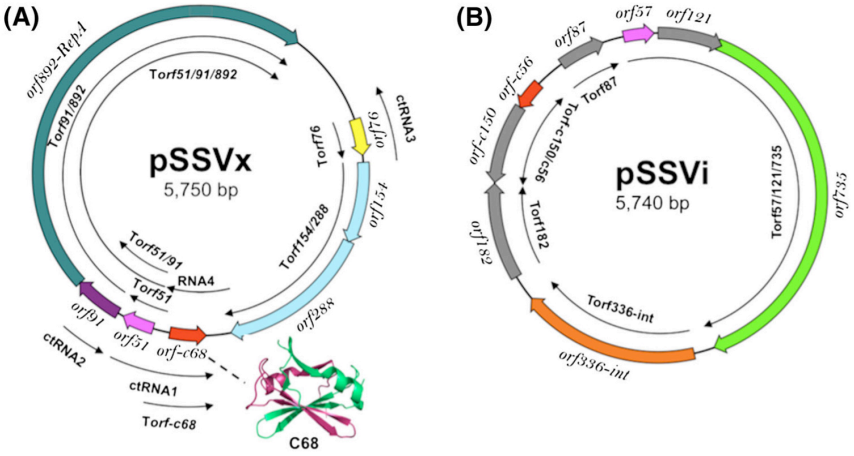


Fig. 5 Genomic map of the satellites pSSVx and pSSVi. ORFs homologues between the two genomes are filled in the same color. The pSSVx transcriptional map has been experimentally determined (by Northern blot and Primer extension) (Contursi et al. 2007, 2010) while that of pSSVi has been deduced from microarray data (Ren et al. 2013). **a** The biggest ORF (*dark green*) encodes for a Primase/Replicase which is homologue to that of the pRN family. The ORFs filled in *light blue* encode for proteins involved in DNA prepackaging and are homologues to two of the 13 ORFs of the *Fuselloviridae* core set. *orf91* and *orf51* encode for putative transcriptional factors, while *orf-c68* and *orf76* encode for experimentally determined transcriptional factors (Contursi et al. 2011 and manuscript submitted). **b** *orf735*, filled in green, encodes for the pSSVi replicase, the putative integrase is encoded by *orf336*, while ORFs in *gray* are orphans

absent from all known pRN plasmids that replicate autonomously.

An additional insight into the formation of satellites viral particles has been gained from another study of *Sulfolobus* genetic elements in which pXZ1, a plasmid that carries an SSV1-type integrase gene and co-exists with SSV4. This plasmid does not have the capability of forming virions but it carries an integrase gene (*int*) that is highly similar to the *int* gene of SSV4; a stretch of 500-bp DNA sequence is identical for the two *int* genes except for one single mismatch at the integration attachment sites, which guides the integration of the plasmid and virus into different tRNA sites in the host genome (Peng 2008).

Together, the genetic elements discussed above, including pSSVx, pSSVi, pXZ1, SSV2 and SSV4, provide very suitable materials to unravel the mechanism of virion packaging for fuselloviruses and their satellites.

Tripartite interactions among SSV2, pSSVi and their host

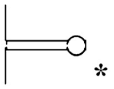
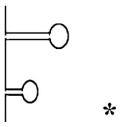
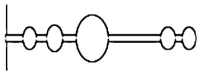
Interactions between pSSVi, SSV2 and the host *S. solfataricus* were investigated by transcriptomic analysis through microarrays. In the pSSVi-containing *S. solfataricus* P2 cells, the integrated form of pSSVi is propagated passively over generations. All pSSVi genes were expressed constitutively at a low level except for the integrase gene, which is partitioned in the host genome. Upon infection with SSV2 (or SSV1), pSSVi replication is activated, producing multiple copies of the episomal DNA on which an intact *int* gene is yielded. Under this condition, expression is elevated strongly for all pSSVi genes. Interestingly, it appears that there is a mutual interaction between the two genetic elements such that the replication of pSSVi facilitates that of its helper virus since the copy number of the latter also increases (Ren et al. 2013).

Notably, elevated expression of pSSVi genes correlates with an enhanced level of the expression from SSV2 early genes. Also the expression level of some host genes is elevated, including that encoding the RNA polymerase subunit F, RpoF. Therefore, it is plausible to suggest that pSSVi stimulates transcription of SSV2 early genes by regulating the expression level of RpoF and/or of other host factors (Ren et al. 2013).

pSSVx: a model for studying gene regulation circuits in *Archaea*

Similar to SSV2, pSSVx exhibits inducible DNA replication in the natural host at a late growth phase. We have demonstrated that the pSSVx genes are differentially expressed during its life cycle, providing a good model for studying regulation of gene expression in *Archaea* (Contursi et al. 2007, 2010). Three putative transcriptional factors seem to be implicated in the regulation of gene expression at the *rep* locus including ORF-c68, the virus satellite-specific factor, and ORF51 and ORF91, the plasmidic DNA-binding proteins (Fig. 5).

Table 2 Features of the pSSVx transcripts as experimentally determined by Northern blot, Primer extension and RT-PCR

Transcript	Transcription start site	Transcription termination sites	Encoded protein	Expressed in the plasmidic phase	Expressed in the viral phase	Promoter feature	Termination signal
T_{orf76}	1095	1325/1405	Transcription factor	Y	N	Canonical	
$T_{orf154/288}$	1367	3091/3150	Proteins involved in DNA packaging	Y	Y	Canonical	ttttttcatttt
T_{orfc68}	3027	2676/2710	Swapped-hairpin-like transcription factor	Y	Y	Canonical	ttatttttccgttattt
T_{orf51}	3150	3343	Putative CopG protein	Y	Y	Atypical	
$T_{orf51/91}$	3150	3550/3612	Putative CopG protein/Putative Zinc-motif protein	Y	Y	Atypical	
$T_{orf91/892}$	3338	669/709	Putative Zinc-motif protein/ RepA	N	Y	Atypical	Tttttttctttt *
$T_{orf51/91/892}$	3150	669/709	Putative CopG protein/Putative Zinc-motif protein/RepA	N	Y	Atypical	tttttttctttt
ctRNA1	3341	2710/2676	X	Y	N	Canonical	ttatttttccgttattt
ctRNA2	3557	3370/3324	X	Y	N	Canonical	ttttttactccatttt
ctRNA3	1383	2023/1056	X	Y	Y	Atypical	tttcgccttttt
RNA4	2686	3091/3150	X	Y	Y	Canonical	ttttttcatttt

The hypothesized DNA-binding sites for the CopG protein are the promoters Pr1 and/or Pr2, whereas ORF-c68 factor binds to its own regulatory sequence (Pr3)

The termination signals marked with an asterisk are predicted

A systematic study on initiation and termination of gene transcription for the pSSVx genome has been conducted. TSSs have been mapped for all the transcripts based on which archaeal canonical promoters as well as atypical promoter sequences have been identified (Table 2). The overlapping arrangement of genes increases the number of proteins that may be produced from a small genome but requires more complex regulatory mechanisms for the modulation of gene expression (see below).

Analyzing the expression of pSSVx genes in the natural host has revealed that they exhibit a temporal regulation. At an early growth phase (plasmidic phase) pSSVx is strictly controlled to a low copy number and transcripts that are specifically detected in this phase, i.e., ORF76, ctRNA1 and ctRNA2, are likely to be important for the stringent control of plasmid replication (Table 2). When pSSVx replication is induced, the expression from the putative replicase operon is strongly elevated and this occurs concurrently with the highest copy number attained by pSSVx (viral phase) (Fig. 6). Furthermore, the expression of three additional transcripts (*Torf51*, *Torf154/orf288* and RNA4) appears to be proportional to pSSVx copy number throughout the entire growth of the host (Contursi et al. 2007, 2010).

Northern analysis of *Torf892-RepA* expression indicates that *Torf51/orf91/orf892-RepA* and *Torf91-orf892-RepA*, the two *rep* transcripts, are only expressed at a later growth phase when virus replication is induced. Since ctRNA1 and ctRNA2 are oppositely transcribed to *orf51* and *orf91* and are highly expressed at an early growth phase of host cells, they are likely to be implicated in the inhibition of the replicase expression at this stage, possibly causing premature termination of *rep* transcripts. Indeed, as revealed from our transcriptional data, transcription terminates at the 3′ ends of *Torf51* and *Torf51-orf91* messengers when ctRNA1 and ctRNA2 are expressed at a consistent high

Extremophiles (2014) 18:473–489

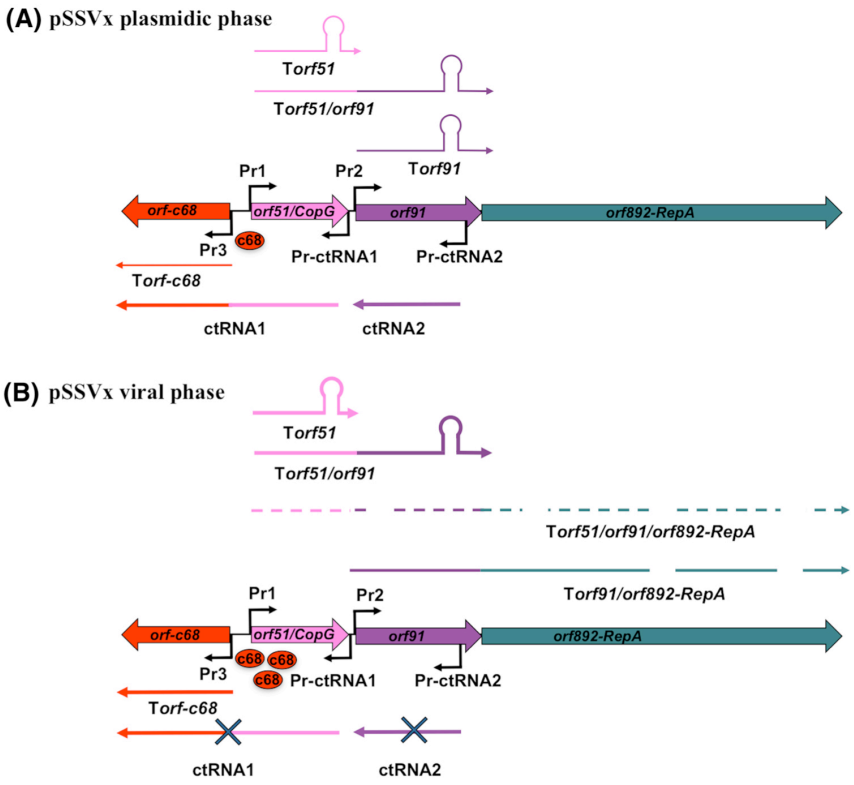


Fig. 6 Schematic representation of the transcription regulation of the pSSVx at the *rep* locus. The transcription of the ORF892-RepA can start at both the promoters Pr1 and Pr2 but with differential initiation frequency as indicated by *solid* (highly transcribed messenger RNAs) and *dashed* (low abundance transcripts) *arrows* in the viral phase of pSSVx replication. The increase in the relative abundance of the other transcripts at the *rep* locus in the passage from the plasmidic to the viral phase is represented by the *thickness* of the *lines*. Differential transcription termination is indicated by the relative length of the transcripts and stem-and-loop structures. ORF892-RepA transcripts are *sketched* as interrupted lines to outline nuclease susceptibility and consequent fast turnover (**b**, viral phase)

level during the plasmidic phase (Fig. 6a). Conversely, at a late growth phase levels of ctRNA1 and ctRNA2 are undetectable (Contursi et al. 2007, 2010) (Fig. 6b). Therefore, the relative abundance of the mRNAs and relatives ctRNAs might move the balance either towards premature termination or to anti-termination. Similarly ctRNA3 may regulate the expression of *orf76*. Regulation via anti-sense RNAs is a common mechanism among bacterial plasmids and anti-sense RNAs are frequently found in archaeal transcriptomes (Siemering et al. 1994; del Solar and Espinosa 2000, Brantl 2002a, b; Tang et al. 2002, 2005; Dodd et al. 2005; Kwong et al. 2006; Straub et al. 2009; Wurtzel et al. 2009).

Besides the shift of the equilibrium from premature termination (more properly a bacterial mode) to anti-termination/read-through (more typically a viral mode) for the *Torf51*, *Torf51/orf91*, *Torf51/orf91/orf892-RepA* messengers, another level of regulation is effective in controlling the expression of the RepA protein at post-transcriptional level and consists in the degradation of the two *rep* messengers (*Torf91/orf892-RepA* and *Torf51/orf91/orf892-RepA*) during the viral phase of pSSVx (Contursi et al. 2007). The degradation of specific transcripts is a typical viral tactic for controlling gene expression at the post-transcriptional and protein synthesis levels, when a compensation is necessary to correct the "sloppiness" in the transcription regulation. This type of mechanism is indeed employed not only by the well known bacterial T4 and lambda phages (Mosig and Hall 1994) but also by the archaeal and closely related fuselloviruses, SSV1 (Reiter et al. 1987b).

The regulatory copy number control systems are usually negative feedback loops that often involve a constitutively expressed regulatory molecule, the abundance of which is determined by gene dosage (directly proportional to plasmid copy number), which negatively affects *rep* gene expression at transcriptional and/or translational levels. The apparent candidate as a "sensor" of the copy number fluctuation of pSSVx is the putative copy number control protein (copG homolog) encoded by *orf51*. The relative abundance of *Torf51* strictly parallels all the variations of the plasmid copy number during the pSSVx life cycle. CopGs generally bind their own promoters (Pr1 in the case of pSSVx) repressing both their own and the transcriptionally coupled expression of the *rep* gene. Biochemical characterization of ORF-c68 has revealed that it specifically binds to a target site upstream of the TATA-box and the BRE element in its own promoter, suggesting that it

functions as a transcriptional activator (Fig. 6a, b). This agrees well with the simultaneous increase of its own transcript as well as of the protein. The biological function and contribution of ORF-c68 to the pSSVx life cycle are still unclear thus far. The messenger *Torf-c68* accumulates even after the plasmid copy number has reached its plateau value, namely up to the very final stage of the pSSVx life cycle. It has been hypothesized that the acquisition of *orf-c68* (as well as of the *orf154* and *orf288*) by the pSSVx genome has been crucial for acquiring the viral nature and the responsiveness to viral stimuli of the helper (Contursi et al. 2011, 2013).

Biochemical characterization of *Sulfolobus* spindle-shaped virus Ragged Hills proteins

Structural and biochemical analyses of *Fuselloviridae* conserved ORFs are functional to unravel the archaeal viral biology. Two proteins encoded by the fusellovirus SSV8 have been characterized, i.e., E73 and D212.

The protein E73 represents an interesting variation on the theme of the Ribbon–Helix–Helix transcription regulators. Indeed, it possesses an additional α-helix located at its C terminal, thus giving rise to the RH3 motif. Likewise most of the RHH proteins, also E73 is an homodimer strengthened by interactions between the N-terminal β-strands of each monomer. The third α-helix stabilizes the quaternary structure by wrapping the helix α2 of the nearby subunit and therefore contributing to the formation of a tightly intertwined dimer (Schlenker et al. 2012; Contursi et al. 2013). E73 shares significant homology with other proteins encoded by other SSV genomes like SSV2 79a, SSV4 73, SSV5 GP23 and SSV6 GP17 and with E51 of SSV1, thus suggesting that this peculiar fold suits with the mechanisms and molecular components required for transcription regulation in the *Fuselloviridae* family.

Orthologs of D212 are present in many *Fuselloviridae* (Wiedenheft et al. 2004). The overall structure of D212 resembles that of type II restriction endonuclease. Furthermore, D212 bears the PD-(D/E)XK catalytic motif that is conserved in this nuclease superfamily. Since, other classes of nucleases involved in DNA replication, repair and recombination also possess the same association of this fold with the PD-(D/E)XK catalytic residues, similar roles also for the SSV8 protein have been hypothesized. A number of functions related to the life cycle and/or viral evolution mechanisms of the *Fuselloviridae* might require the catalytic activity of such nuclease. For instance, D212 might: (1) recognize DNA branch point of the SSV8 genome with a specific geometric structure and specific sequences or (2) participate to the correction of replication defects or other physical damages to the viral genome or (3) catalyze the excision of a covalent closed circular DNA virus from a concatemer as suggested by Redder et al. 2009 (Menon et al. 2010). Although, none of these in vivo roles have been proven, a hint about its function comes from a study performed on the SSV1 homolog (D212). Indeed, the host growth slowed down in cells infected with the mutated virus SSV-Δd244. This effect is probably due to defects in SSV-Δd244 replication or in the resolution of its replication intermediates that might lead to accumulation of aberrant DNA and in turn to a slowing down of the host growth (Iverson and Stedman 2012).

SSV9 integrates not only at tRNA locations

SSV9 was isolated from a hot (75 °C) acidic (pH 4.0) pool in the Geyser Valley region of a National park in the Kamchatka peninsula (Russia) (Wiedenheft et al. 2004).

Analysis of potential *attP* and *attA* sites for directing integration of the viral genome, pointed out to 4 putative *attA* sites on the *S. solfataricus* genome, one at an aspartic acid tRNA, the other two at the glutamic acid tRNA genes (E1 and E2) and the fourth at a non-tRNA site. Differently from SSV8, SSV1 and SSV2 that integrate at single site, SSV9 is able to target all the four predicted *attA* sites. Nevertheless, an exact copy of the tRNA gene is reconstituted only upon integration at the glutamic acid gene, whereas integration at the remaining locations is not as precise (Wiedenheft et al. 2004). This is the first report of an integrating virus that target a non-tRNA site (Campbell 1992; Reiter et al. 1989) and the biological significance of multiple integration events needs to be further investigated.

Shuttle vectors based on *Fuselloviridae* genomes

Genetic analysis is fundamental in dissecting the molecular biology and the physiology of any organisms. Effective and more sophisticated "hot" expression systems are indispensable not only for better understanding gene function and protein sorting in vivo but also for the production of fully active and best performing enzymes.

Despite the extensive characterization of extrachromosomal elements in *Sulfolobus (*Zillig et al. 1996; Prato et al. 2006; Snyder et al. 2003; Lipps 2006; Fu and Johnson 2012; Prangishvili 2013), only few of them have been used as template for the setting up of suitable genetic tools (Atomi et al. 2012; Zheng et al. 2012; Zhang et al. 2013). The putative replication origin of the fusellovirus SSV1 has been employed for the construction of the first shuttle vector for *S. solfataricus*, the pEXSs, in which the hygromycin phosphotransferase gene was the selective marker (Cannio et al. 1998, 2001; Contursi et al. 2004). Although it was able to replicate efficiently both in *S. solfataricus* and *E. coli*, the low copy number in the hyperthermophilic host hindered its applicability for heterologous expression

or genetic manipulations (Fiorentino et al. 2009; Contursi et al. 2003).

Later on, shuttle vectors based on the whole SSV1 genome were developed in a first instance for studying viral ORFs essentiality and subsequently used successfully for heterologous gene expression and complementation (Stedman et al. 1999; Jonuscheit et al. 2003; Albers et al. 2006; Iverson and Stedman 2012). The vector pKMSD48 contains the pBluescript II SK + *E. coli* plasmid site specifically inserted into e178, an ORF encoded by the polycistronic messenger T_5. Despite its bigger size (about 18 kbp), if compared to the wild-type SSV1 (15.5 Kbp), pKMSD48 is efficiently packaged into viable virions and able to spread through *Sulfolobus* cultures without undergoing to recombination or rearrangement events (Stedman et al. 1999). A similar shuttle vector (pJM03) has been assembled by fusing the SSV1 moiety with exogenous sequences within the non-essential ORF e51, i.e., the *E. coli* plasmid pUC18, the *lacS* expression cassette and the *pyrE/F* genes as reporter and selection marker, respectively (Jonuscheit et al. 2003). Its efficient spreading as well as the successful genetic complementation showed by this vector, prompted its use for heterologous expression at high level of hyperthermophilic proteins (Albers et al. 2006).

Another suitable shuttle vector for *S. solfataricus* was obtained by fusing site specifically the pSSVx chromosome with an *E. coli* plasmid replicon. The resulting recombinant vector was able to propagate in *E. coli* under ampicillin selection and at high copy number in *S. solfataricus* with no recombination/integration events occurring. The stable maintenance of this vector was achieved by inserting the *lacS* cassette as selection marker. The resulting recombinant vector relied on the co-presence of the SSV2 for its replication, packaging and spreading (Aucelli et al. 2006). Moreover, it was successfully used for homologous gene transfer and overexpression of the *sso1354* gene encoding one of the three putative endo-β-1-4-glucanases from *S. solfataricus* (Limauro et al. 2001; Girfoglio et al. 2012).

Concluding remarks

Fuselloviridae provide unique models for studying molecular virology and molecular biology in *Archaea*. To date, SSV1 and SSV2 are still the only archaeal viruses exhibiting inducible virus production. These unique features have facilitated their research ever since their discovery. Early studies on constitutive and UV-inducible transcription represent a milestone in archaeal transcription study including definition of archaeal basal promoter elements as well as revealing signals of archaeal transcription termination (Reiter et al. 1988a, b). Then, an extensive study of the pSSVx transcription has revealed complex pre- and post-transcriptional mechanisms to regulate the expression of its replicase during the physiological induction of pSSVx replication in the presence of SSV2 (Contursi et al. 2007, 2010). Transcriptomic analysis of viral and host genomes using whole-genome microarrays has unraveled cascade regulation of gene expression. However, very limited insights have been gained into the regulation of these processes and we still know very little about the life cycle of fuselloviruses and their regulation.

One of the first outstanding questions for the fuselloviral research is what are the main regulators leading to the inducible virus production? Identification of the SSV1 lysogenic regulator represents an important step towards the mechanism of the induction process (Fusco et al. 2013) but detailed mechanisms remain to be investigated. Crystal structures have been solved for several putative SSV1 transcriptional factors. A combinatorial approach of biochemical and genetic assays will be more powerful to study functions of these putative transcriptional factors such that their target sites will be revealed from biochemical analysis whereas insights into their functions will be gained from sophisticated genetic analyses involving gene knockout and mutagenesis.

Both viruses exhibit a cascade regulation of gene expression during virus induction, a process that very likely involves host transcriptional factors. Fortunately, tools for genetic study are not only available for SSV1 and SSV2/pSSVx (Aucelli et al. 2006; Iverson and Stedman 2012) but also for their hosts (She et al. 2009; Wagner et al. 2009). Therefore, SSV1 and SSV2/pSSVx provide excellent model systems for these investigations.

There are several important themes in fuselloviral research to be investigated including mechanisms of viral DNA replication, cell receptors for virus infection, viral and host components that participate in virion assembly and cellular extrusion. Mechanism of virus satellites formation is a very exciting theme as it reveals how archaeal plasmids could have exploited the fuselloviruses to enhance their possibilities for survival. Finally, the arm race between archaeal hosts and their viruses involves a novel genetic entity termed CRISPR systems, i.e., "Clustered regularly interspaced palindromic repeats" (CRISPR) and CRISPR-associated proteins (Garrett et al. 2011; Held and Whitaker 2009). Apparently fuselloviruses have developed the capability of circumventing the CRISPR systems of their hosts because they infect these archaeal organisms despite of the fact that these hosts carry many significant matches to fuselloviral genomes identified in their CRISPR loci (Redder et al. 2009). Clearly, focused researches are required to provide important insights into these research themes.

Acknowledgments Researches in the authors' laboratories are supported by the grant "Programma F.A.R.O, IV° tornata", founded by Università Federico II di Napoli and "Compagnia di San Paolo" (Naples Laboratory) and grants (DFF–1323-00330 and FTP/11-106683) from Danish Council of Independent Research (Copenhagen laboratory). We thank Peter Redder and Li Huang for providing pictures of fuselloviruses and Prof. Simonetta Bartolucci for valuable scientific discussions throughout the preparation of the manuscript.

References

Ackermann HW, Prangishvili D (2012) Prokaryote viruses studied by electron microscopy. Arch Virol 157:1843–1849

Albers SV, Jonuscheit M et al (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Appl Environ Microbiol 72:102–111

Albers SV, Forterre P, Prangishvili D, Schleper C (2013) The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. Nat Rev Microbiol 11:713–719

Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. Nat Struct Biol 8:710–714

Argos P, Landy A et al (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J 5:433–440

Arnold HP, She Q et al (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. Mol Microbiol 34:217–226

Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM (2012) Global network of specific virus-host interactions in hypersaline environments. Environ Microbiol 14:426–440

Atomi H, Imanaka T, Fukui T (2012) Overview of the genetic tools in the *Archaea*. Front Microbiol 3:337

Aucelli T, Contursi P, Girfoglio M, Rossi M, Cannio R (2006) A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*. Nucleic Acids Res 34:e114

Bamford DH (2003) Do viruses form lineages across different domains of life? Res Microbiol 154:231–236

Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? Curr Opin Struct Biol 15:655–663

Bartolucci S, Contursi P, Fiorentino G, Limauro D, Pedone E (2013) Responding to toxic compounds: a genomic and functional overview of Archaea. Front Biosci 18:165–189

Bize A, Peng X et al (2008) Viruses in acidic geothermal environments of the Kamchatka Peninsula. Res Microbiol 159:358–366

Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M (2012) Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. J Virol 86:5562–5573

Brantl S (2002a) Antisense RNAs in plasmids: control of replication and maintenance. Plasmid 48:165–173

Brantl S (2002b) Antisense-RNA regulation and RNA interference. Biochim Biophys Acta 1575:15–25

Campbell AM (1992) Chromosomal insertion sites for phages and plasmids. J Bacteriol 174:7495

Cannio R, Contursi P, Rossi M, Bartolucci S (1998) An autonomously replicating transforming vector for *Sulfolobus solfataricus*. J Bacteriol 180:3237–3240

Cannio R, Contursi P, Rossi M, Bartolucci S (2001) Thermoadaptation of a mesophilichygromycin B phosphotransferase by directed evolution in hyperthermophilic Archaea: selection of a stable genetic marker for DNA transfer into *Sulfolobus solfataricus*. Extremophiles 5:153–159

Ceballos RM, Marceau CD, Marceau JO, Morris S, Clore AJ, Stedman KM (2012) Differential virus host-ranges of the *Fuselloviridae* of hyperthermophilic Archaea: implications for evolution in extreme environments. Front Microbiol 3:295

Clore AJ, Stedman KM (2007) The SSV1 viral integrase is not essential. Virology 361:103–111

Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, Bartolucci S (2003) Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. FEMS Microbiol Lett 218:115–120

Contursi P, Pisani FM, Grigoriev A, Cannio R, Bartolucci S, Rossi M (2004) Identification and autonomous replication capability of a chromosomal replication origin from the archaeon *Sulfolobus solfataricus*. Extremophiles 8:385–391

Contursi P, Jensen S, Aucelli T, Rossi M, Bartolucci S, She Q (2006) Characterization of the *Sulfolobus* host-SSV2 virus interaction. Extremophiles 10:615–627

Contursi P, Cannio R, Prato S, She Q, Rossi M, Bartolucci S (2007) Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. J Bacteriol 189:6339–6350

Contursi P, Cannio R, She Q (2010) Transcription termination in the plasmid/virus hybrid pSSVx from *Sulfolobus islandicus*. Extremophiles 14:453–463

Contursi P, D'Ambrosio K et al (2011) C68 from the *Sulfolobus islandicus* plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. Biochem J 435:157–166

Contursi P, Fusco S, Limauro D, Fiorentino G (2013) Host and viral transcriptional regulators in *Sulfolobus*: an overview. Extremophiles 17:881–895

Cortez D, Forterre P, Gribaldo S (2009) A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. Genome Biol 10:R65

del Solar G, Espinosa M (2000) Plasmid copy number control: an ever-growing story. Mol Microbiol 37:492–500

Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G (2009) Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in archaea by MarR family proteins. J Mol Biol 388:559–569

Dodd IB, Shearwin KE, Egan JB (2005) Revisited gene regulation in bacteriophage lambda. Curr Opin Genet Dev 15:145–152

Eilers BJ, Young MJ, Lawrence CM (2012) The structure of an archaeal viral integrase reveals an evolutionarily conserved catalytic core yet supports a mechanism of DNA cleavage in trans. J Virol 86:8309–8313

Fiorentino G, Ronca R, Bartolucci S (2009) A novel *E. coli* biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein. Appl Microbiol Biotechnol 82:67–77

Fiorentino G, Del Giudice I, Bartolucci S, Durante L, Martino L, Del Vecchio P (2011) Identification and physicochemical characterization of BldR2 from *Sulfolobus solfataricus*, a novel archaeal member of the MarR transcription factor family. Biochemistry 50:6607–6621

Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5–16

Forterre P, Prangishvili D (2013) The major role of viruses in cellular evolution: facts and hypotheses. Curr Opin Virol 3:558–565

Fröls S, Gordon PM, Panlilioc MA, Schleper C, Sensen CW (2007) Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology 365:48–59

Fu CY, Johnson JE (2012) Structure and cell biology of archaeal virus STIV. Curr Opin Virol 2:122–127

Fusco S, She Q, Bartolucci S, Contursi P (2013) T(lys), a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. J Virol 87:5926–5936

Garrett RA, Shah SA, Vestergaard G, Deng L, Gudbergsdottir S, Kenchappa CS, Erdmann S, She Q (2011) CRISPR-based immune systems of the Sulfolobales: complexity and diversity. Biochem Soc Trans 39:51–57

Gibrat JF, Madej T, Bryant SH (1996) Surprising similarities in structure comparison. Curr Opin Struct Biol 6:377–385

Girfoglio M, Rossi M, Cannio R (2012) Cellulose degradation by *Sulfolobus solfataricus* requires a cell-anchored endo-beta-1-4-glucanase. J Bacteriol 194:5091–5100

Grummt I, Maier U, Ohrlein A, Hassouna N, Bachellerie JP (1985) Transcription of mouse rDNA terminates downstream of the 3′ end of 28 s rRNA and involves interaction of factors with repeated sequences in the 3′ spacer. Cell 43:801–810

Guixa-Boixareu N, Calderon-Paz JI, Heldal M, Bratbak G, Pedros-Alio C (1996) Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. Aquat Microb Ecol 11:215–227

Guo X, Huang L (2010) A superfamily 3 DNA helicase encoded by plasmid pSSVi from the hyperthermophilic archaeon *Sulfolobus solfataricus* unwinds DNA as a higher-order oligomer and interacts with host primase. J Bacteriol 192:1853–1864

Guo L, Brugger K, Liu C et al (2011) Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. J Bacteriol 193:1672–1680

Häring M, Rachel R, Peng X, Garrett RA, Prangishvili D (2005) Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. J Virol 79:9904–9911

Held NL, Whitaker RJ (2009) Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. Environ Microbiol 11:457–466

Helmer-Citterich M, Anceschi MM, Banner DW, Cesareni G (1988) Control of ColE1 replication: low affinity specific binding of Rop (Rom) to RNAI and RNAII. EMBO J 7:557–566

Huet J, Schnabel R, Sentenac A, Zillig W (1983) Archaebacteria and eukaryotes possess DNA dependent. RNA polymerases of a common type. EMBO J 2:1291–1294

Iverson E, Stedman K (2012) A genetic study of SSV1, the prototypical fusellovirus. Front Microbiol 3:1–7

Jaubert C, Danioux C, Oberto J, Cortez D, Bize A, Krupovic M, She Q, Forterre P, Prangishvili D, G Sezonov (2013) Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. Open Biol 3(4):130010. doi:10.1098/rsob.130010

Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C (2003) A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. Mol Microbiol 48:1241–1252

Keeling PJ, Klenk HP et al (1996) Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. Plasmid 35:141–144

Keeling PJ, Klenk HP et al (1998) *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. Extremophiles 2:391–393

Kletzin A, Lieke A, Urich T, Charlebois RL, Sensen CW (1999) Molecular analysis of pDL10 from *Acidianus ambivalens* reveals a family of related plasmids from extremely thermophilic and acidophilic archaea. Genetics 152:1307–1314

Koonin EV (1992) Archaebacterial virus SSV1 encodes a putative DnaA-like protein. Nucleic Acids Res 11:1143

Koonin EV, Senkevich TG, Dolja VV (2006) The ancient Virus World and evolution of cells. Biol Direct 19:1–29

Kraft P, Kümmel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM (2004a) Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol 78:7438–7442

Kraft P, Oeckinghaus A et al (2004b) Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. J Virol 78:11544–11550

Krupovic M, Prangishvili D, Hendrix RW, Bamford DH (2011) Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol Mol Biol Rev 75:610–613

Krupovic M, White MF, Forterre P, Prangishvili D (2012) Postcards from the edge: structural genomics of archaeal viruses. Adv Virus Res 82:33–62

Kwong SM, Skurray RA, Firth N (2006) Replication control of staphylococcal multiresistance plasmid pSK41: an antisense RNA mediates dual-level regulation of Rep expression. J Bacteriol 188:4404–4412

Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, Douglas T, Young MJ (2009) Structural and functional studies of archaeal viruses. J Biol Chem 284:12599–12603

Le Romancer M, Gaillard M, Geslin C, Prieur D (2007) Viruses in extreme environments. Extremophiles 6:17–31

Limauro D, Cannio R, Fiorentino G, Rossi M, Bartolucci S (2001) Identification and molecular characterization of an endoglucanase gene, celS, from the extremely thermophilic archaeon *Sulfolobus solfataricus*. Extremophiles 5:213–219

Lipps G (2006) Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus*. Extremophiles 10:17–28

Lipps G, Rother S, Hart C, Krauss G (2003) A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J 22:2516–2525

Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W (1984) SAV 1, a temperate UV.-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J 3:2165–2168

Menon SK, Maaty WS et al (2008) Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology 376:270–278

Menon SK, Eilers BJ, Young MJ, Lawrence CM (2010) The crystal structure of D212 from *Sulfolobus* spindle-shaped virus ragged hills reveals a new member of the PD-(D/E)XK nuclease superfamily. J Virol 84:5890–5897

Mochizuki T, Yoshida T, Tanaka R, Forterre P, Sako Y, Prangishvili D (2010) Diversity of viruses of the hyperthermophilic archaeal genus Aeropyrum, and isolation of the Aeropyrum pernix bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. Virology 402:347–354

Mochizuki T, Krupovic M, Pehau-Arnaudet G, Sako Y, Forterre P, Prangishvili D (2012) Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. Proc Natl Acad Sci USA 109:13386–13391

Mosig G, Hall DH (1994) Regulation of gene expression. In: Karam JD (ed) Molecular biology of bacteriophage T4. ASM Press, Washington, DC, pp 127–193

Muskhelishvili G, Palm P, Zillig W (1993) SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. Mol Gen Genet 237:334–342

Nadal M, Mirambeau G, Forterre P, Reiter W-D, Duguet M (1986) Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature 321:256–258

Okutan E, Deng L, Mirlashari S, Uldahl K, Halim M, Liu C, Garrett RA, She Q, Peng X (2013) Novel insights into gene regulation of

the rudivirus SIRV2 infecting *Sulfolobus* cells. RNA Biol 10:875–885
Oren A, Bratbak G, Hendal M (1997) Occurrence of virus-like particles in the Dead Sea. Extremophiles 1:143–149
Ortmann AC, Wiedenheft B, Douglas T, Young M (2006) Hot crenarchaeal viruses reveal deep evolutionary connections. Nat Rev Microbiol 4:520–528
Ortmann AC, Brumfield SK et al (2008) Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus* turreted icosahedral virus. J Virol 82:4874–4883
Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology 185:242–250
Peng X (2008) Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of *Sulfolobus* and the implications for plasmid survival. Microbiology 154:383–391
Peng X, Holz I, Zillig W, Garrett RA, She Q (2000) Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. J Mol Biol 303:449–454
Peng X, Garrett RA, She Q (2012) Archaeal viruses–novel, diverse and enigmatic. Sci China Life Sci 55:422–433
Pietilä MK, Roine E, Paulin L, Kalkkinen N, Bamford DH (2009) An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. Mol Microbiol 72:307–319
Pietilä MK, Laurinmäki P, Russell DA, Ko CC, Jacobs-Sera D, Butcher SJ, Bamford DH, Hendrix RW (2013) Insights into head-tailed viruses infecting extremely halophilic archaea. J Virol 87:3248–3260
Pina M, Bize A, Forterre P, Prangishvili D (2011) The archeoviruses. FEMS Microbiol Rev 35:1035–1054
Porter K, Russ BE, Dyall-Smith ML (2007) Virus–host interactions in salt lakes. Curr Opin Microbiol 10:418–424
Porter K, Tang SL, Chen CP, Chiang PW, Hong MJ, Dyall-Smith M (2013) PH1: an archaeovirus of Haloarcula hispanica related to SH1 and HHIV-2. Archaea 2013:456318
Prangishvili D (2003) Evolutionary insights from studies on viruses of hyperthermophilic archaea. Res Microbiol 154:289–294
Prangishvili D (2013) The wonderful world of archaeal viruses. Annu Rev Microbiol 8:565–585
Prangishvili D, Garrett RA (2004) Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. Biochem Soc Trans 32:204–208
Prangishvili D, Garrett RA (2005) Viruses of hyperthermophilic Crenarchaea. Trends Microbiol 13:535–542
Prangishvili D, Stedman K, Zillig W (2001) Viruses of the extremely thermophilic archaeon *Sulfolobus*. Trends Microbiol 9:39–43
Prangishvili D, Forterre P, Garrett RA (2006a) Viruses of the Archaea: a unifying view. Nat Rev Microbiol 4:837–848
Prangishvili D, Garrett RA, Koonin EV (2006b) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res 117:52–67
Prato S, Cannio R, Klenk HP, Contursi P, Rossi M, Bartolucci S (2006) pIT3, a cryptic plasmid isolated from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* IT3. Plasmid 56:35–45
Quax TE, Voet M, Sismeiro O, Dillies MA, Jagla B, Coppée JY, Sezonov G, Forterre P, van der Oost J, Lavigne R, Prangishvili D (2013) Massive activation of archaeal defense genes during viral infection. J Virol 87:8419–8428
Quemin ER, Lucas S, Daum B, Quax TE, Kühlbrandt W, Forterre P, Albers SV, Prangishvili D, Krupovic M (2013) First insights into the entry process of hyperthermophilic archaeal viruses. J Virol 87:13379–13385
Rachel R, Bettstetter M, Hedlund BP, Häring M, Kessler A, Stetter KO, Prangishvili D (2002) Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. Arch Virol 147:2419–2429
Redder P, Peng X, Brügger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ Microbiol 11:2849–2862 John Wiley & Sons Ltd
Reiter WD, Palm P et al (1987a) Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. Mol Gen Genet 206:144–153
Reiter WD, Palm P, Yeats S, Zillig W (1987b) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol Gen Genet 209:270–275
Reiter WD, Palm P, Zillig W (1988a) Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res 16:1–19
Reiter WD, Palm P, Zillig W (1988b) Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. Nucleic Acids Res 16:2445–2459
Reiter WD, Palm P, Yeats S (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res 17:1907–1914
Ren Y, She Q, Huang L (2013) Transcriptomic analysis of the SSV2 infection of *Sulfolobus solfataricus* with and without the integrative plasmid pSSVi. Virology 441:126–134
Rice G, Stedman K et al (2001) Viruses from extreme thermal environments. Proc Natl Acad Sci USA 98:13341–13345
Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakić M, Young MJ, Lawrence CM, Copié V (2012) Structural studies of E73 from a hyperthermophilic archaeal virus identify the "RH3" domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. Biochemistry 51:2899–2910
Schleper C, Kubo K, Zillig W (1992) The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sci USA 89:7645–7649
Serre MC, Letzelter C, Garel JR, Duguet M (2002) Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. J Biol Chem 277:16758–16767
She Q, Peng X, Zillig W, Garrett RA (2001a) Gene capture in archaeal chromosomes. Nature 409:478
She Q, Singh RK et al (2001b) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc Natl Acad Sci USA 98:7835–7840
She Q, Brügger K, Chen L (2002) Archaeal integrative genetic elements and their impact on genome evolution. Res Microbiol 153:325–332
She Q, Shen B, Chen L (2004) Archaeal integrases and mechanisms of gene capture. Biochem Soc Trans 32:222–226
She Q, Zhang C, Deng L, Peng N, Chen Z, Liang YX (2009) Genetic analyses in the hyperthermophilic archaeon *Sulfolobus islandicus*. Biochem Soc Trans 37:92–96
Siemering KR, Praszkier J, Pittard AJ (1994) Mechanism of binding of the antisense and target RNAs involved in the regulation of IncB plasmid replication. J Bacteriol 176:2677–2688
Sime-Ngando T, Lucas S, Robin A, Tucker KP, Colombet J, Bettarel Y, Desmond E, Gribaldo S, Forterre P, Breitbart M, Prangishvili D (2011) Diversity of virus-host systems in hypersaline Lake Retba, Senegal. Environ Microbiol 13: 1956–1972

Snyder JC, Young MJ (2011) Advances in understanding archaea-virus interactions in controlled and natural environments. Curr Opin Microbiol 14:497–503

Snyder JC, Stedman K, Rice G, Wiedenheft B, Spuhler J, Young MJ (2003) Viruses of hyperthermophilic Archaea. Res Microbiol 154:474–482

Stedman KM, Schleper C, Rumpf E, Zillig W (1999) Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. Genetics 152:1397–1405

Stedman KM, She Q et al (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: sSV1 and SSV2. Res Microbiol 154:295–302

Stedman KM, Kosmicki NR, Diemer GS (2013) Codon usage frequency of RNA virus genomes from high-temperature acidic-environment metagenomes. J Virol 87:1919

Straub J, Brenneis M, Jellen-Ritter A, Heyer R, Soppa J, Marchfelder A (2009) Small RNAs in haloarchaea: identification, differential expression and biological function. RNA Biol 6:281–292

Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. Proc Natl Acad Sci USA 99:7536–7541

Tang TH, Polacek N, Zywicki M et al (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol Microbiol 55:469–481

Wagner M, Berkner S, Ajon M, Driessen AJM, Albers SV (2009) Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. Biochem Soc Trans 37: 97–101

Wang Y, Duan Z, Zhu H, Guo X, Wang Z, Zhou J, She Q, Huang L (2007) A novel *Sulfolobus* non-conjugative extrachromosomal genetic element capable of integration into the host genome and spreading in the presence of a fusellovirus. Virology 363: 124–133 Elsevier Publishers

Wiedenheft B, Stedman K et al (2004) Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol 78:1954–1961

Woese CR, Magrum LJ, Fox GE (1978) Archaebacteria. J Mol Evol 11:245–251

Wu RY, Zhang RG et al (2003) Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. J Biol Chem 278: 20240–20244

Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R (2009) A single-base resolution map of an archaeal transcriptome. Genome Res 20:133–141

Yeats S, McWilliam P, Zillig W (1982) A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. EMBO J 1:1035–1038

Zhan Z, Ouyang S, Liang W, Zhang Z, Liu ZJ, Huang L (2012) Structural and functional characterization of the C-terminal catalytic domain of SSV1 integrase. Acta Crystallogr D Biol Crystallogr 68:659–670

Zhang C, Cooper TE, Krause DJ, Whitaker RJ (2013) Augmenting the genetic toolbox for *Sulfolobus islandicus* with a stringent positive selectable marker for agmatine prototrophy. Appl Environ Microbiol 79:5539–5549

Zheng T, Huang Q, Zhang C, Ni J, She Q, Shen Y (2012) Development of a simvastatin selection marker for a hyperthermophilicacidophile, *Sulfolobus islandicus*. Appl Environ Microbiol 78:568–574

Zillig W, Palm P, Reiter WD, Gropp F, Puhler G, Klenk HP (1988) Comparative evaluation of gene expression in archaebacteria. Eur J Biochem 173:473–482

Zillig W, Kletzin A et al (1994) Screening for *Sulfolobales*, Their Plasmids and Their Viruses in Icelandic Solfataras. Syst Appl Microbiol 16:609–628

Zillig W, Prangishvilli D et al (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic *Archaea*. FEMS Microbiol Rev 18:225–236

1.5 Aims of the project

All bioprocesses that use microorganisms to produce value-added products are potentially subjected to great losses of the transforming biomasses caused by naturally occurring viral infections. The arms race among microorganisms and viruses has led to a continuous evolution (on the host side) of efficient mechanism to cope with infections and, in turn, to attempts (on the viral side) to defeat the host defences. Of course, microorganisms evolve towards the improvement of their fitness rather than in favor of the biotechnological efficiency. As results, in order to achieve safeguarded bioprocesses, natural defense systems have to be engineered to generate infection-insensitive microorganisms. Otherwise, natural selection of such immune-boosted strains would take a very long time.
The model organism used in this project is *S. solfataricus* P2, i.e., a well-characterized strain isolated from the sulphuric hot spring of Pisciarelli, near Naples [33]. As mentioned above, the CRISPR-Cas response of this archaeon is very complex, due to the co-presence of the type I and III systems and the abundance of CRISPR loci [25]. Consequently, *S. solfataricus* offers a valuable opportunity to study the activity of the CRISPR-Cas system in a heterogeneous context. As viral models, two fuselloviruses have been chosen, i.e., SSV1 and SSV2. Interestingly enough, even though these two viruses show morphological and genetic similarities, they display very different life cycles. SSV1 was originally isolated from its natural host *Sulfolobus shibatae* strain B12 (Beppu, Japan), but is able to infect *S. solfataricus* as well. It is a temperate virus that maintains a low copy number in both the hosts and establishes a harmonious consistence with the infected cell until its replication is induced by UV-light exposure [34]. On the other hand, the replication induction of SSV2 is regulated by the physiological state of the host cell and rather than by an external stimulus as for SSV1. Once infected the native host *S. isandicus* REY15/4 isolated from a sulfuric hot spring in Reykjanes (Iceland), the SSV2 copy number is kept constant and low (around 1-3 copies per cell) until the host enters into the stationary phase of growth when a steep increase of the copy number occurs (25-50 copies per cell) [34]. Noteworthy, this effect has not been observed for the permissive host *S. solfataricus* P2, in which SSV2 shows a copy number comparable to that of the induce state of the natural host throughout the growth of *S. solfataricus*. This led to hypothesize that a transcription factor encoded by the natural host *S. islandicus* (and not by *S. solfataricus*) was likely involved in the regulation of the viral replication induction in the natural host.

Thanks to the combination of similarities and differences, these two viruses are suitable to investigate the *S. solfataricus* response against infections, particularly as concerns the CRISPR-Cas systems activation. The main aim of this project is to study virus-host and virus-virus relationships, in order to gain insight useful to lay the bases for the setting up of procedures to generate Virus Insensitive Thermophilic Strains (VITSs).

References

[1] Labrie SJ, Samson JE, Moineau S. (2010). Bacteriophage resistance mechanisms. *Nat Rev Microbiol*. **8**:317-327.

[2] Stern A, Sorek R. (2011). The phage-host arms race: shaping the evolution of microbes. *Bioessays*. **33**:43-51.

[3] Samson JE, Magadán AH, Sabri M, Moineau S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol*. **11**:675-687.

[4] Van Valen L. (1973). A new evolutionary law. *Evol Theor*. **1**:1-30.

[5] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**:1709-1712.

[6] Makarova KS, Wolf YI, Koonin EV. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res*. **41**:4360-4377.

[7] Sturino JM, Klaenhammer TR. (2006). Engineered bacteriophage-defence systems in bioprocessing. *Nat Rev Microbiol*. **4**:395-404.

[8] Sorek R, Kunin V, Hugenholtz P. (2008). CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol*. **6**:181-186.

[9] Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of *Archaea*, *Bacteria* and mitochondria. *Mol Microbiol*. **36**:244-246.

[10] Jansen R, Embden JD, Gaastra W, Schouls LM. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. **43**:1565-1575.

[11] Marraffini LA, Sontheimer EJ. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet*. **11**:181-190.

[12] Sorek R, Lawrence CM, Wiedenheft B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem*. **82**:237-266.

[13] Heler R, Marraffini LA, Bikard D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol Microbiol*. **93**:1-9.

[14] van der Oost J, Westra ER, Jackson RN, Wiedenheft B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol*. **12**:479-492.

[15] Marraffini LA, Sontheimer EJ. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*. **463**:568-571.

[16] Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. **9**:467-477.

[17] Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet*. **26**:335-340.

[18] Bikard D, Marraffini LA. (2013). Control of gene expression by CRISPR-Cas systems. *F1000Prime Rep*. **5**:47.

[19] Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, Garrett RA. (2010). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol*. **79**:35-49.

[20] Dyall-Smith M. (2011). Dangerous weapons: a cautionary tale of CRISPR defence. *Mol Microbiol*. **79**:3-6.

[21] Deveau H, Garneau JE, Moineau S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol*. **64**:475-493.

[22] Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. **1**:7.

[23] Ghosh S. (2015). A CRISPR cut to bacteria. *Nat Struct Mol Biol*. **22**:36.

[24] Song CW, Lee J, Lee SY. (2015). Genome engineering and gene expression control for bacterial strain development. *Biotechnol J*. **10**:56-68.

[25] Zhang J, White MF. (2013). Hot and crispy: CRISPR-Cas systems in the hyperthermophile *Sulfolobus solfataricus*. *Biochem Soc Trans*. **41**:1422-1426.

[26] Manica A, Schleper C. (2013). CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus *Sulfolobus*. *RNA Biol*. **10**:671-678.

[27] Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, Garrett RA. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol Microbiol*. **72**:259-272.

[28] Erdmann S, Garrett RA. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol*. **85**:1044-1056.

[29] Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, Sdano M, Peng N, She Q, Copié V, Young MJ, White MF, Lawrence CM. (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J Biol Chem*. **286**:21643-21656.

[30] Manica A, Zebec Z, Steinkellner J, Schleper C. (2013). Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res*. **41**:10509-50917.

[31] Prangishvili D. (2013). The wonderful world of archaeal viruses. *Annu Rev Microbiol*. **67**:565-585.

[32] Krupovic M, Quemin ER, Bamford DH, Forterre P, Prangishvili D. (2014). Unification of the globally distributed spindle-shaped viruses of the *Archaea*. *J Virol*. **88**:2354-2358.

[33] Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I. (1980). The *Sulfolobus*-"*Caldariella*" Group: Taxonomy on the Basis of the Structure of DNA-Dependent RNA Polymerases. *Arch. Microbiol*. **125**:259-269.

[34] Contursi P, Fusco S, Cannio R, She Q. (2014). Molecular biology of fuselloviruses and their satellites. *Extremophiles.* In press. DOI 10.1007/s00792-014-0634-0.

Chapter II

– The fusellovirus SSV1 establishes a harmonious co-existence with the host *Sulfolobus solfataricus* –

2.1 Summary

Sulfolobus spindle-shaped virus 1 (SSV1) is the founding member of the *Fuselloviridae* family, which compresses to date 10 viruses (SSV1-SSV9 and ASV1) that generally exhibit a spindle-shaped capsid of ca. 60 by 100 nm. Notably, this unique virion morphology seems to be a hallmark of viruses infecting *Archaea*, since it has never been observed for bacterial and eukaryal viruses. For the past three decades, SSV1 has served as a model for studying host-virus interactions as well as for developing genetic tools. It is the best-characterized member of *Fuselloviridae* and is the only one showing an UV-inducible life cycle. SSV1 exhibits a genomic region similar to that of the bacteriophage lambda, whose lytic cycle is induced by UV-light exposure as well. Upon infecting a host cell, this fusellovirus integrates one copy of its genome into the host chromosome at an arginyl-tRNA gene, thus forming a provirus. However, unlike the well-characterized lambda phage, for which lysogenic cells harbour only the provirus, SSV1 lysogens carry also some copies of the episomal DNA. This has led to the hypothesis that SSV1 expresses a minimal set of genes to ensure a basal level of replication that is required for maintaining its carrier stage. Moreover, structural proteins (VP1, VP2 and VP3) have been shown to be constitutively expressed under conditions in which viral replication is not induced, thus allowing the production of viral particles by lysogenic cells. Upon exposure to UV light, SSV1 exhibits a temporally coordinated pattern of gene expression. At first, it activates the expression of a UV-inducible transcript (T_{ind}), followed by the transcription of the early (T_5, T_6 and T_9), late ($T_{1/2}$, T_3, T_x, and $T_{4/7}$) and late-extended ($T_{4/7/8}$) RNAs. This cascade of events leads, in turn, to the induction of the SSV1 genome replication and eventually to a steep increase of the viral titre. However, since the SSV1 genome encodes for a number of quasi-orphan proteins that do not have detectable homologues in the databases, it was possible only to speculate about their functions basing on their time of appearance in the aftermath of the UV irradiation. This has led, in turn, to the necessity of performing structural and functional analyses to unravel their functions. For instance, the structure of several SSV1 transcription factors (TFs) has been solved revealing that, despite the lack of homology, most of these viral TFs are bacterial like. Nevertheless, the lack of functional characterization of these proteins has limited the dissection of fundamental processes, such as: i) the virion uptake, assembly and release, ii) the transcriptional regulation, iii) the genome replication and iv) the switch from lysogeny to induction.

Intriguingly, unless the host is stressed by the exogenous stimulus (i.e. the UV-light exposure), SSV1 seems to establish a harmonious co-existence with the cell and does not induce the activation of the CRISPR-Cas system in *S. solfataricus* (see Chapter III). With the aim of investigating how SSV1 can maintain this stable relationship with the host, experiments were carried out to dissect molecular mechanisms underpinning the transition from the lysogenic to the induced state of SSV1. As model host it was chosen *S. solfataricus* P2, because its genome was completely sequenced and microarrays were available for whole-transcriptome analysis (see Chapter III). Indeed, preliminary microarray data showed that a region of the SSV1 genome located nearby the T_{ind} transcript, and previously considered not transcribed was, indeed, actively expressed during the carrier stage. This led to the identification of a novel uncharacterized transcript, named T_{lys}, which encodes for a 6.3-kDa protein, termed F55. This latter shows sequence identity with negative regulators that fold into the ribbon-helix-helix DNA-binding motif (RHH motif), thus suggesting that it might function as transcription repressor. DNA band-shift assays

demonstrated that F55 is indeed a DNA-binding protein that is able to site-specifically recognize target sequences in the promoters of the early-induced T_5, T_6, and T_{ind} transcripts, as well as in its own promoter. Interestingly, the strongest affinity was observed with promoters of T_5 and T_6, and an apparent cooperativity was observed for the binding to that of T_{ind}. Moreover, these target sequences encompass both the transcription start sites (TSSs) and the B recognition elements (BREs) of the corresponding promoters. These results suggested that, likewise the lambda repressor CI, F55 may be the key regulator of the lysogenic/induction switch (for further details see par. 2.2).

Nevertheless, its activity needed to be studied *in vivo* in the aftermath of the UV irradiation. However, in peer reviewed published literature, it was not clear how the fluence (or UV dose) and the irradiance have been measured. Indeed, to date, the fluence (J m^{-2}) administered to the cells has been only empirically determined and no attention has been paid to monitor the irradiance (J m^{-2} s^{-1}). Furthermore, the negative effect on the host viability, that the UV treatment implies, has been so far underestimated. For these reasons, it was resolved to set up an irradiation protocol based on instrumentally measured parameters. By doing so, it was not only demonstrated a clear dose-response relationship between the strength of the UV irradiation and the host survival percentage, but also that cells lethality is significantly reduced by tuning the irradiance. In particular, a fluence of 45 J m^{-2} combined with a mild irradiance of 0.5 J m^{-2} s^{-1}, turned out to be suitable to preserve host viability and, in turn, to obtain a high increase of the viral copy number as well as of the viral titre (viral particles per milliliter of culture). Hence, it seemed the best condition in which the role of F55 could be investigated (for further details see par. 2.3).

Even though F55 is the only putative transcription repressor for which a defined role in the regulation of a fuselloviral life cycle has been proposed, an *in vivo* survey was sorely needed to investigate molecular events occurring at the UV-inducible region of the SSV1 genome upon irradiation. Several lines of evidence indicated that F55 is indeed the key regulator of the lysogeny/induction transition of SSV1, i.e.: i) F55 is the only transcriptional regulator expressed by SSV1 in the lysogenic state; ii) it binds *in vivo* to the promoters of T_{ind}, T_{lys}, T_5 and T_6 in the absence of UV stimulus; iii) the expression of the UV-inducible transcript T_{ind} is shut down when F55 binds to its promoter, and iv) transcription activation of T_{ind} occurs only upon releasing of F55 in the immediate aftermath of the UV irradiation. In particular, it was demonstrated that F55 dissociates upon UV irradiation first from the T_{ind} (2 hours post-treatment) and subsequently from the T_5 and T_6 promoters (4 hours post-treatment). This is in a perfect agreement with its differential affinity towards these regulatory sequences *in vitro* ($T_5 = T_6 > T_{ind}$). Intriguingly, the transcription of T_5 and T_6 is repressed during the SSV1 carrier stage and is activated in UV-irradiated cells four hours post-treatment. These data strongly indicate that F55 regulates the transcriptional activity of these mRNAs in a fashion similar to that of T_{ind}, i.e. by repressing their expression during the lysogenic growth and allowing their transcription activation through clearance of their target sites. Altogether these data demonstrate that F55, likewise the CI repressor of lambda, acts as the key switch regulator involved in the lysogeny/induction transition of SSV1. However, some differences between lambda and SSV1 have been revealed. For example, lambda lysogens harbour the viral DNA exclusively as provirus and the only viral gene expressed is the *cI*. The lysogenic/lytic cycle is strictly regulated since viral progeny is produced only upon UV irradiation and causes cell lysis. By contrast, in SSV1 lysogens the provirus co-exists together with some episomal copies and a

constitutive extrusion of viral particles occurs. Therefore, the lysogeny of SSV1 could be better defined as a carrier stage. Indeed, the UV irradiation only enhances the rate of SSV1 replication and progeny extrusion without causing cell lysis. This difference is mirrored by the expression regulation of F55, which is not as stringent as for the functional homolog CI of lambda. From an evolutionary point of view, this kind of host-virus relationship, which is typical for all *Fuselloviridae*, could be co-evolved as consequence of the non-lytic nature of these viruses (for further details see par. 2.4).

2.2 T_{lys}, a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at promoters of the early genes

JVI Journals.ASM.org

T_{lys}, a Newly Identified *Sulfolobus* Spindle-Shaped Virus 1 Transcript Expressed in the Lysogenic State, Encodes a DNA-Binding Protein Interacting at the Promoters of the Early Genes

Salvatore Fusco,[a] Qunxin She,[b] Simonetta Bartolucci,[a] Patrizia Contursi[a]

Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Naples, Italy[a]; Danish Archaea Centre, Department of Biology, University of Copenhagen, Copenhagen N, Denmark[b]

While studying the gene expression of the *Sulfolobus* spindle-shaped virus 1 (SSV1) in *Sulfolobus solfataricus* lysogenic cells, a novel viral transcript (T_{lys}) was identified. Transcriptional analysis revealed that T_{lys} is expressed only in the absence of UV irradiation and is downregulated during the growth of the lysogenic host. The correponding gene *f55* lies between two transcriptional units (T_6 and T_{ind}) that are upregulated upon UV irradiation. The open reading frame *f55* encodes a 6.3-kDa protein which shows sequence identity with negative regulators that fold into the ribbon-helix-helix DNA-binding motif. DNA-binding assays demonstrated that the recombinant F55, purified from *Escherichia coli*, is indeed a putative transcription factor able to recognize site specifically target sequences in the promoters of the early induced T_5, T_6, and T_{ind} transcripts, as well as of its own promoter. Binding sites of F55 are included within a tandem-repeated sequence overlapping the transcription start sites and/or the B recognition element of the pertinent genes. The strongest binding was observed with the promoters of T_5 and T_6, and an apparent cooperativity in binding was observed with the T_{ind} promoter. Taking together the transcriptional analysis data and the biochemical evidences, we surmise that the protein F55 is involved in the regulation of the lysogenic state of SSV1.

Sulfolobus shibatae B12 isolated from a hot spring in Beppu, Japan (1), is the natural host of the *Sulfolobus* spindle-shaped virus 1 (SSV1). This archaeal virus represents a prototype of nine such viruses belonging to the *Fuselloviridae* family (2, 3). Most of them exhibit a spindle-shaped capsid of ca. 60 by 100 nm with a short tail at one end, but extended morphotypes have also been observed for those with bigger genomes (3). For the past several decades, SSV1 has served as a model for studying archaeal viruses as well as for developing genetic tools (4–7). The envelope of SSV1 virions is composed of three coat proteins (VP1, VP2, and VP3), among which VP2 is directly associated with viral DNA, whereas VP1 and VP3 form the surface structure of the viral particles (8).

Several putative transcription factors encoded by the SSV1 genome, include (i) E51 and C80, which are ribbon-helix-helix proteins (RHH) belonging to the CopG family, and (ii) A45, A79, and B129 containing C_2H_2 zinc finger-like motifs (9). Four putative DNA-binding proteins were structurally and/or functionally characterized, i.e., (i) the fold of D63 resembles that of the "repressor of primer" (ROP) (10), an adaptor protein that regulates colE1 plasmid copy number in *Escherichia coli* (11), (ii) F63 (12) and F112 (13) harbor a winged helix fold, which is a typical feature of proteins that belong to the SlyA (14) and MarR (15) subfamilies of winged-helix DNA-binding proteins, and (iii) E73 which displays the "RH3" domain, a variant of the RHH motif (16). Possibly, these putative transcription factors could play a function in the cascade regulation of SSV1 gene expression induced by UV irradiation. However, it is very unlikely that any of them could have a role in controlling the switch between the viral lysogenic versus UV-inducible life cycle, since they are not expressed during the lysogenic growth (unpublished data).

SSV1 has a narrow host range, i.e., among the three *Sulfolobus* model organisms, including *S. solfataricus*, *S. acidocaldarius*, and *S. islandicus* (17, 18), this virus only infects *S. solfataricus* strains isolated from the solfataric field of Pisciarelli near Naples (Italy) (19).

Moreover, SSV1 is the only known crenarchaeal virus that exhibits a genomic region similar to that of the bacteriophage lambda and shows an UV-inducible life cycle (20). Instead, the closely related SSV2 shows a growth-phase-dependent induction of its replication (21–23).

Upon infection, the double-stranded SSV1 genome of 15,465 bp integrates into the host chromosome within an arginyl-tRNA gene via the integrase D355 encoded by the viral genome (17). Since attachment sites are located within the *d355* coding sequence, the integration results in partitioning of this gene that inactivates the enzyme, leading to gene capture events in the host chromosome (24, 25).

A previous analysis of the SSV1 transcriptome revealed a well-coordinated temporal expression of nine transcripts in the native host after UV irradiation (26). Subsequently, a more detailed characterization of the UV induction process was conducted using microarray analysis, which largely confirmed the previous results. Upon UV irradiation, the expression of SSV1 genome starts from an UV-inducible immediate-early transcript (T_{ind}, which is located between the two "back-to-back" oriented early transcripts T_5 and T_6). This is then followed by the synthesis of the early (T_5, T_6, and T_9), the late ($T_{1/2}$, T_3, T_x, and $T_{4/7}$) transcripts and finally of the late- extended polycistronic messenger ($T_{4/7/8}$). These transcripts are organized in the viral genome according to the chrono-

Received 15 February 2013 Accepted 6 March 2013
Published ahead of print 20 March 2013
Address correspondence to Patrizia Contursi, contursi@unina.it.

doi:10.1128/JVI.00458-13

logical order of their expression; this fashion of regulation is reminiscent of that used by many bacteriophages and eukaryotic viruses (27). The first insight into archaeal transcription was gained by mapping minimal promoter elements and transcription terminators of SSV1 (28, 29).

Despite the extensive characterization of the viral expression upon UV exposure, the molecular components and the mechanisms underpinning the maintenance of the SSV1 lysogenic state are still poorly understood.

We recently investigated the regulation of gene expression in the lysogenic state of this virus by microarrays using as a host SSV1-InF1, an infected *S. solfataricus* P2-derived strain (30). It has been revealed that a region of the SSV1 genome located nearby the T_{ind} transcript, previously considered not transcribed (26), was indeed actively expressed. This led to the identification of a novel uncharacterized transcript, named T_{lys}, which encodes for a 6.3-kDa protein, termed F55. In the present study, the analysis of T_{lys} transcription, as well as the functional characterization of the protein F55, is reported. The results of these analyses provide further insights into the SSV1 life cycle and suggest that F55 might be the regulator of the lysogenic state of this fusellovirus.

MATERIALS AND METHODS

Strains, media, and growth conditions. *S. solfataricus* InF1, a uracil auxotrophic mutant isolated previously (30), was the host for generating an SSV1 lysogenic strain. In brief, 1 to 2 μl of the *S. shibatae* B12 supernatant containing SSV1 virions was spotted onto the soft layer of a Gelrite plate seeded with uninfected cells of InF1. After 2 to 3 days of incubation at 75°C, turbid halos (plaques) appeared on the plate surface as result of an infection-dependent inhibition of host growth. SSV1-infected cells were extracted from plaques and revitalized in liquid medium. Thereafter, a single colony, herein named SSV1-InF1, was isolated by Gelrite plating and purified by restreaking onto plate for three times. Liquid cultures of *S. solfataricus* were grown aerobically in TYSU medium, a glycine-buffered Brock's basal salt solution, supplemented with 0.1% tryptone, 0.05% yeast extract, 0.2% sucrose (wt/vol), and 0.02 mg of uracil/ml; the pH was finally adjusted to 3.2 with concentrated H_2SO_4. InF1 and SSV1-InF1 samples from frozen cultures were inoculated into 50 ml of TYSU medium, and culture incubation was conducted in 100- or 250-ml Erlenmeyer flasks with a long neck using an Innova 3100 water bath shaker (New Brunswick Scientific Corp.). The incubation temperature was 75°C with a shaking rate of 150 rpm, and cell growth was monitored spectrophotometrically at 600 nm. When cultures reached the logarithmic phase of growth, they were diluted to an optical density at 600 nm (OD_{600}) of 0.05 in 50 ml of fresh medium and let to grow up to an OD_{600} value of 0.3 to 0.5. Subsequently, cultures were diluted again down to 0.05 OD_{600} and split in three parallel growing cultures that were harvested at three different OD_{600} values: 0.4 (exponential phase), 0.8 (late exponential phase), and 1.2 (early stationary phase). Samples were centrifuged in 50-ml Falcon tubes at 3,000 × *g* for 10 min using the Centrifuge 5810R (Eppendorf), and pellets were treated for total DNA (using DNeasy tissue kit; Qiagen) and RNA (using TRIzol; Sigma reagent) preparations.

qPCR for determination of SSV1 genome copy number. Two primer couples were designed using Primer3 software, available at the website (http://frodo.wi.mit.edu/), in order to amplify: (i) a 155-bp fragment of the SSV1 single-copy gene *vp2* (*vp2*-fw, 5′-TATAAATTGTTATAGACATAGAACGCTGTA-3′; *vp2*-rv, 5′-TTAAATACTTCTTGTGCCGATAGTCC-3′) and (ii) a 108-bp region of the host single-copy gene *orc1* (*orc1*-fw, 5′-GGAGGGTACATCGCTACCTTATGA-3′; *orc1*-rv, 5′-CAGTAGGGCTGACAGTAAACTACG-3′).

Real-time qPCR amplifications were carried out by means of an iQ5 multicolor real-time PCR detection system (Bio-Rad). The QuantiFast SYBR Green PCR kit (Qiagen) was used for the preparation of the real-time PCR mixtures. In brief, 12.5 μl of 2× QuantiFast SYBR green PCR buffer and primers at 1 μM concentration were premixed and dispensed into a 96-well plate (Thermowell Gold PCR Plates). Subsequently, appropriated dilutions of standards and samples were added in RNase-free H_2O to reach a final volume of 25 μl. The thermal cycling protocol was as follows: an initial denaturation step of 5 min at 95°C, followed by 35 cycles of 40 s at 95°C, 40 s at 62°C, and 40 s at 72°C. The fluorescence signal was measured at the end of each extension step. A final step at 72°C has been carried out for 10 min at the end of the 35th cycle. iQ5 optical system software uses the threshold cycle (C_T) value of each amplified sample to calculate the initial template copy number by means of standard curves. Total DNA samples extracted from InF1 cultures were used as negative control for *vp2* amplification.

The setting up of suitable experimental conditions for the absolute quantification was achieved by (i) testing the specificity of the amplified products by melting curves and gel electrophoresis analyses and (ii) constructing standard curves for *orc1* and *vp2* amplicons (data not shown) (31). For this purpose, 10-fold dilutions ranging from 10^8 to 10^2 molecules per μl were used. The C_T value, which is defined as the cycle number at which the fluorescence generated within a reaction crosses the fluorescence threshold, was measured in duplicate for each dilution, and standard curves were constructed by plotting C_T values against the initial copy number of molecules.

The absolute copy number determination of *orc1* and *vp2* was obtained from two independent experiments, where each experimental point (OD_{600} values of 0.4, 0.8, and 1.2) was analyzed at least in triplicate. The SSV1 genome copy number per host cell was calculated dividing the total copy number of *vp2* by the total copy number of *orc1*. A biological confirmation of the data was achieved by testing the SSV1 genome copy number in total DNA samples derived from a parallel culture of SSV1-InF1.

Northern blot analysis. Total RNA samples (20 μg) from three different growth phases (see above) were run on a denaturing, formaldehyde-containing 2.0% (wt/vol) agarose gel and then transferred onto a nylon membrane (Hybond-XL; Amersham-Pharmacia). T4 polynucleotide kinase (Fermentas Life Sciences) was used to label 5′ ends of single-stranded oligonucleotides T_{lys}rv (5′-AAGTTCTTCAATGCGTCTTCTGATT-3′) and T_{ind}fw (5′-TCTGAGCTACTAATACTGCTTGAAT-3′) with radioactive [γ-^{32}]ATP, according to the manufacturer's instructions. The primers *Sso2359*fw (5′-AGATGAATGGGTTAATGTT-3′) and *Sso2359*rv (5′-CACTAAAACATAAATATCCC-3′) were used in PCR amplifications of the *Sso2359*-specific probe from total DNA of InF1. Therefore, the amplicon was 5′-end radiolabeled, purified, and used for hybridization in normalization experiments. The purification of radiolabeled oligonucleotides was achieved by gel filtration chromatography using illustra Nick columns (Amersham Biosciences). Hybridization with single- and double-stranded DNA probes was carried out as described elsewhere (22). The probes were eventually removed by 10 min of boiling in 0.1% sodium dodecyl sulfate (SDS) to reuse the membrane for subsequent hybridizations. The relative abundance of T_{lys} and of the housekeeping *Sso2359* transcripts was evaluated by quantifying the radioactive signals using a Molecular Dynamics Bio-Rad PhosphorImager (Quantity One software). The size of the T_{lys} mRNA was determined using the RNA molecular weight markers (Roche) as standards.

Primer extension analysis. Total RNA (20 μg) and 1 pmol (10^5 cpm) of 5′-labeled primer T_{lys}rv (5′-AAGTTCTTCAATGCGTCTTCTGATTG-3′) were coprecipitated by using 2.5 volumes of 96% ethanol for 30 min at −80°C. After centrifugation, the pellets were resuspended in the reverse transcription buffer purchased from Ambion and denatured for 3 min at 65°C, frozen in dry ice, and thawed on ice. Reactions were incubated for 30 min at 37°C for the annealing, and then deoxynucleoside triphosphates (final concentration, 2 mM each) and 20 U of RNase inhibitor (Promega) was added. The extension of the primer was performed using 8 U of Moloney murine leukemia virus reverse transcriptase (M-MLV RV; Ambion) for 1 h at 48°C. Sequencing reactions of the corresponding DNA

TABLE 1 Sequences of oligonucleotides used in band-shift assays

Name	Promoter region	Oligonucleotide		
		Strand[a]	Sequence (5′–3′)	Length (nt)
T_6-D-TR	T_6	PS	GATATATAGATAGAGTATAGATAGAGTAAAG	31
T_6-D-TR	T_6	MS	CTTTACTCTATCTATACTCTATCTATATATC	31
T_{lys}-G-TR	T_{lys}	PS	GTCATATTAATATAGTATAGATAGCGTATGC	31
T_{lys}-G-TR	T_{lys}	MS	GCATACGCTATCTATACTATATTAATATGAC	31
T_{ind}-E-TR	T_{ind}	MS	GTTGTATAAGATACATAAGATACACAG	27
T_{ind}-E-TR	T_{ind}	MS	CTGTGTATCTTATGTATCTTATACAAC	27
T_5-A-SR	T_5	MS	GATTTATAGATAGAGTGGGA	20
T_5-A-SR	T_5	MS	TCCCACTCTATCTATAAATC	20

[a] PS and MS indicate plus and minus strands, respectively.

region, to be used as reference ladders, were carried out with the same primer, T_{lys}rv, and the fmol DNA cycle sequencing system kit (Promega) according to the manufacturer's instructions. The resulting products were separated on denaturing 6% polyacrylamide gels, along with the DNA sequencing reaction products.

Bioinformatic analysis. An open reading frame (ORF) encoding for a 55-amino-acid protein, named F55, was identified on the T_{lys} transcript with the NCBI ORF finder software and the ExPASy translate tool. Similarity comparison of F55 was performed using BLAST (Basic Local Alignment Search Tool). The workbench Jalview 2 (32) was used for multiple sequence alignment comparison and secondary structure prediction. Regions upstream of the transcription start site (TSS) of T_{lys} and downstream of the stop codon were checked in order to identify canonical elements of archaeal promoters and terminators, respectively.

Cloning of the ORF *f55* and overexpression and purification of the recombinant protein. The coding region *f55* was amplified from the SSV1 genome using the primers *f55*-fw (5′-TATAGTATAGATAGCATATGCC GAGGAA-3′) and *f55*-rv (5′-GAGAGATAGAG TATACTCGAGCATTT AAG-3′), in which restriction sites are underlined. The MinElute PCR purification kit (Qiagen) was used to purify the PCR product according to the manufacturer's instructions. After digestion with NdeI and XhoI (Roche), the amplicon was ligated to the NdeI/XhoI-digested pET30(b), yielding pET-*f55*. The cloned fragment was sequenced to verify its identity. Overexpression of F55 was carried out at 0.6 OD_{600} in *E. coli* BL21-CodonPlus (DE3)-RIL cells upon addition of 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 3 h. Cells from 1 liter of culture were harvested by centrifugation, and pellets were resuspended in 20 ml of lysis buffer, 10 mM Tris-HCl (pH 8.0), containing Complete 600 protease inhibitor cocktail tablets (Roche). The cells were lysed by sonication for 5 min, alternating 20 s of pulse-on and 20 s of pulse-off, by means of an ultrasonic liquid processor (Heat System Ultrasonic, Inc.). Lysates were centrifuged at 30,000 × *g* (SW41 rotor; Beckman) for 30 min to clarify the crude extracts. In order to purify the recombinant F55, the lysates were dialyzed overnight against buffer A (50 mM Tris-HCl, 200 mM NaCl [pH 7.0]) and then loaded onto a 1-ml cation-exchange Resource S column (GE Healthcare) connected to a fast-performance liquid chromatography system (AKTA). After a washing step with the buffer A, the elution was performed with a linear salt gradient from 200 to 800 mM NaCl. Protein-containing fractions were collected and analyzed by SDS-PAGE to detect the protein F55 (~6.3 kDa). The F55-containing fractions were pooled, dialyzed against buffer A, and loaded onto a Superdex 75 16/60 column (GE Healthcare) for gel filtration chromatography. To determine the native molecular mass of the protein, the purified F55 at different concentrations (0.5 and 2.0 mg/ml) was applied in a volume of 200 μl to an analytical Superdex PC75 column (3.2 by 30 cm) connected to an AKTA Explorer system (GE Healthcare) equilibrated with buffer A at a flow rate of 0.04 ml/min. The column was calibrated using a set of gel filtration markers (low range; GE Healthcare), including ovalbumin (43.0 kDa), chymotrypsinogen A (25.0 kDa), RNase A (13.7 kDa), and aprotinin (6.5 kDa).

Electrophoresis mobility shift assay (EMSA). Oligonucleotides used in band-shift assays (Table 1) were annealed as follows. Equimolar amounts of plus (PS) and minus (MS) strands were mixed in annealing buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA [pH 8.0]), denatured at 95°C for 5 min, and slowly allowed to cool down at room temperature. Double-stranded DNA probes were 5′-end radiolabeled by means of a T4 polynucleotide kinase (Fermentas) in the presence of [γ-^{32}]ATP. Purification of double-stranded probes was performed as described above. Thermal preincubation of the purified protein F55 was conducted for 15 min at 50°C in assay buffer (25 mM Tris-HCl, 50 mM KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol [pH 7.0], 5% glycerol [vol/vol]) supplemented with 500 ng of salmon testes DNA (Sigma) as nonspecific competitor. The labeled probes were added to the binding reactions at a concentration of ~0.03 μM (~ 20,000 cpm). Binding assays were performed with increasing amounts of F55 (from 1 to 24 μM) for 30 min at 50°C. In the displacement experiments, the binding reactions were performed with 8, 10, or 12 μM F55 with the concurrent addition to the EMSA mixtures of increasing amounts of specific, unlabeled probe (i.e., 1:20, 1:50, 1:100, 1:200, 1:300, 1:400, and 1:500 ratios of labeled/unlabeled specific DNA) or salmon testes DNA as nonspecific competitor (1:50, 1:250, 1:500, 1:750, 1:1,000, 1:1,500, 1:2,000, 1:2,500, and 1:3,000 ratios of labeled specific/nonspecific DNA). Finally, samples were loaded onto a 10% polyacrylamide gel prepared in 0.5× Tris-borate-EDTA and run at 10 mA for ~1 h. Gels were transferred onto filter paper, dried, and revealed by using a Molecular Dynamics Bio-Rad PhosphorImager and/or autoradiography. Shifted signals were quantified by means of the QuantityOne software, and dissociation constants (K_d) were calculated using Prism6 (GraphPad Software) by plotting protein amounts against the percentages of shifted signals (Hill Plot). The K_d is defined as the protein concentration at which 50% of the target sequence is bound by the protein.

Nucleotide and amino acid sequences. Nucleotide and protein sequence data reported are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession number TPA BK008732.

RESULTS

Physiological characterization of a *S. solfataricus* SSV1 lysogen. Upon UV irradiation of an SSV1 lysogen, virus DNA replication and virion production are strongly enhanced without apparent lysis of host cells as for most other archaeal viruses (32). The virus is able to rapidly regain lysogenic status in the absence of the UV stimulus (17), indicating that SSV1 has developed a stringent control over the lysogenic growth. To obtain an insight into the virus life cycle, *S. solfataricus* InF1 was infected with SSV1 to yield SSV1-harboring strains. One such strain (SSV1-InF1) was used for further investigations, along with a virus-free culture (InF1). Both cultures grew in a similar fashion (Fig. 1), indicating the establish-

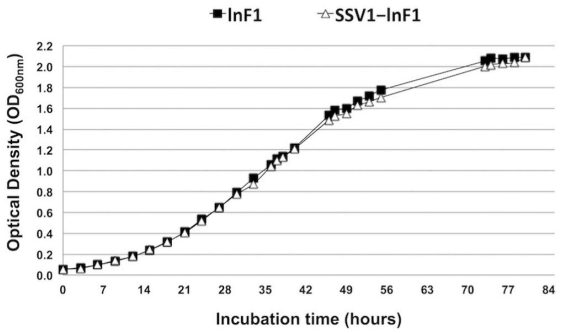


FIG 1 Growth curves of InF1 and SSV1-InF1 strains. The OD_{600} values were measured and plotted versus the incubation time. Both strains grow exponentially until an OD_{600} value of 1.0 and enter into the stationary phase around the 50th hour of incubation. Total RNAs used in the transcriptional analysis (see the Northern blot and primer extension experiments described in Fig. 3) were isolated from cultures harvested at OD_{600} values of 0.4, 0.8, and 1.2.

ment of a harmonious coexistence between host and virus in the lysogenic state.

The copy number of SSV1 genome in the SSV1-InF1 lysogen was estimated by quantitative PCR (qPCR) at different growth phases (OD_{600} values of 0.4, 0.8, and 1.2), through quantification of the amounts of *vp2* and *orc1* genes in the total DNA samples. Since these two genes are present as single copies on viral and host genomes, respectively (24, 33), the ratio between the amount of viral and host genes represents the copy number of SSV1. We found about five copies of SSV1 per chromosome, and this copy number was kept constant during the entire growth (Table 2). Therefore, the variation in transcripts level described below is independent from gene dosage (copy number).

Identification of a novel transcript highly expressed only in the SSV1 lysogenic cells. To gain an insight into the maintenance of the SSV1 lysogenic state, total RNAs were prepared from SSV1-InF1 and InF1 cultures and used for whole-genome microarray analyses. A few viral genes were found to be expressed in the SSV1-harboring cells. These genes could be divided into two categories: (i) those expressed during the entire lysogenic growth, including the ones encoding the structural proteins VP1 and VP3, the integrase D335, A291, and C124, and (ii) those expressed only in the exponential growth phase, i.e., E96, A82, C84/A92, and the capsid-associated VP2 (data not shown).

Whereas the UV-inducible T_{ind} transcript was never detected by Northern blotting and microarray analyses in the lysogenic state (data not shown), a probe derived from the intergenic region of SSV1 located between the promoters of T_6 and T_{ind} (Fig. 2A) showed the presence of a highly expressed transcript, named T_{lys} ("lys" stands for lysogeny), which has never been identified in previous analyses (26, 27). To determine its size, total RNAs were analyzed by Northern blotting with a 5′-radiolabeled single-stranded DNA probe, matching against the T_{lys} transcript. The probe detected a transcript of about 0.25 to 0.30 kb which was constantly expressed during the lysogenic state, albeit its abundance decreased to 20% of the initial amount at a late growth phase (Fig. 3A).

Analyzing the sequence of the intergenic region led to the identification of an ORF (*f55*), which is oriented in the opposite direction with respect to the ORF *b49* lying on the UV-inducible transcript T_{ind} (27) (Fig. 2B). The name *f55* was chosen according to the annotation method by Palm et al. (33). Putative canonical promoter elements, i.e., TATA box and B recognition element (BRE) (28), as well as transcription terminators (23, 29), were identified upstream and downstream of *f55*, respectively (Fig. 2B).

Since the detected mRNA was ~100 nucleotides (nt) longer than the ORF *f55* (168 nt), we determined by primer extension its transcription start site (TSS), which mapped to a T residue located 10 bp upstream of the putative start codon (Fig. 2B). Transcription termination signals are located between 80 and 120 bp downstream of the stop codon. Therefore, the deduced size of T_{lys} was ~300 nt, which is in good agreement with the results of Northern analysis. It must be pointed out the T_{lys} was undetectable upon UV irradiation (26), indicating that its role is specifically related to the lysogenic state of SSV1.

T_{lys} encodes a putative RHH transcription factor that binds to its own promoter. The F55 protein exhibits 40 to 50% sequence identity to several transcription regulators containing the RHH motif and belonging either to the NikR (nickel responsive) (34) or CopG (35) families (Fig. 4).This suggested that F55 could function as a transcription factor as well.

To study this, the gene was cloned into pET30(b) giving pET-*f55* with which the recombinant F55 was expressed in *E. coli* BL21-CodonPlus (DE3)-RIL cells. After purification to homogeneity via two-step procedure, i.e., cation-exchange and gel filtration chromatographies, the purified protein was analyzed by SDS-PAGE. A single band with an apparent molecular mass of ~6.3 kDa appeared (Fig. 5A); this is in good agreement with its predicted molecular mass (6,261 Da). Analytic size-exclusion chromatography of the recombinant protein revealed a size of ~12.4 kDa, indicating that F55 forms a dimer in solution (Fig. 5B). The protein was also found to form dimers in a multi-angle light scattering analysis (data not shown) using a MiniDAWN Treos light-scattering system (Wyatt Technology) as previously described (36).

In the first attempt to study the propensity of F55 to specifically bind to DNA, EMSAs were performed with a 93-bp double-stranded DNA fragment of the predicted promoter region of T_{lys}. The purified protein retarded the migration of the labeled DNA substrate (data not shown), suggesting that the promoter of T_{lys} contains binding motifs recognized by F55. An imperfect tandem-repeated sequence of 22 nt, herein named T_{lys}-G-TR (TR, i.e., tandem repeat), was found overlapping the TSS of T_{lys} (Fig. 6A). The labeled DNA fragment was shifted up when incubated with

TABLE 2 SSV1 copy number estimation by absolute quantification method

Culture[a]	SSV1 copy no. (copies/cell)		
	OD_{600} 0.4	OD_{600} 0.8	OD_{600} 1.2
1	6.1	6.7	5.4
2	5.4	4.8	5.3
3	4.6	4.9	6.5
Avg ± SD (n = 3)	5.4 ± 0.61	5.5 ± 0.87	5.8 ± 0.54

[a] In rows 2 and 3 are listed the results obtained from two independent qPCR experiments on the same total DNA samples, whereas row 3 corresponds to results from another culture of the SSV1-InF1 strain (biological confirmation of the data). The average along with the corresponding standard deviation (SD) is reported in the last row.

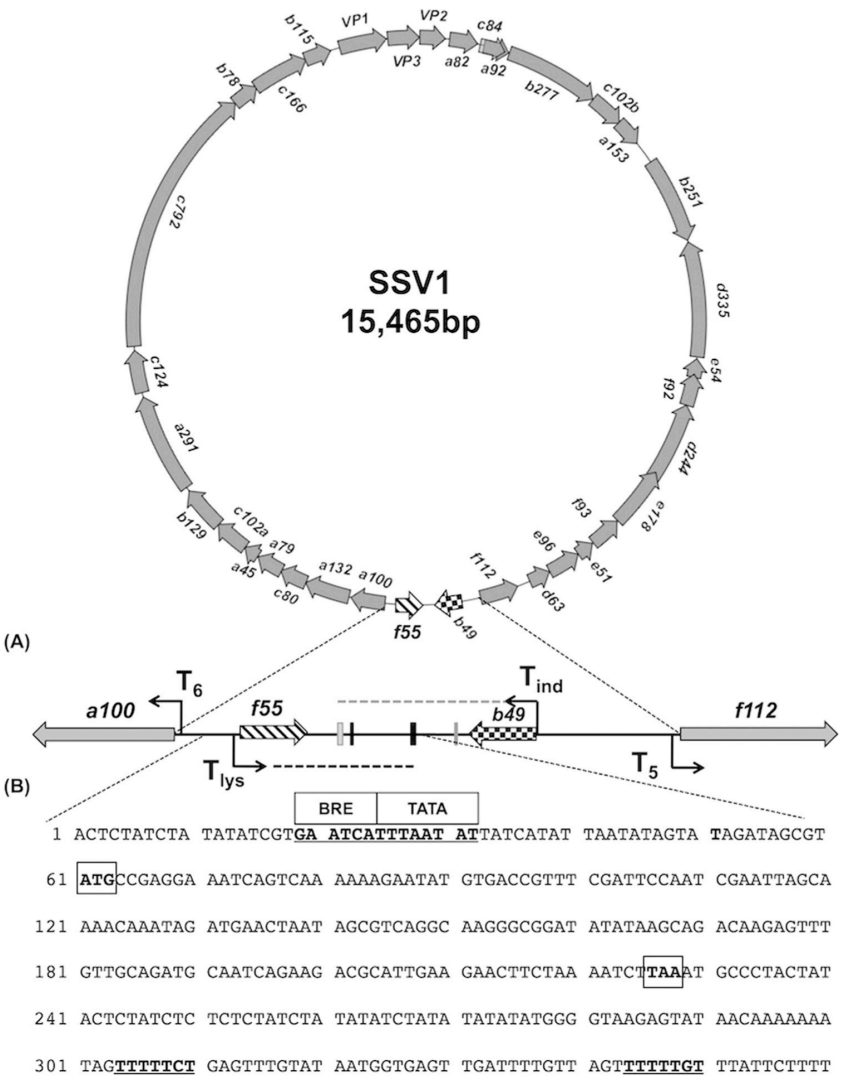


FIG 2 Scheme of the genomic region critical for the UV induction of SSV1. (A) The SSV1 genome map is schematized. ORFs lying on the plus and minus strands are clockwise and anticlockwise oriented. The head-to-head oriented ORFs *b49* (T_{ind}) and *f55* (T_{lys}) are checkered and striped, respectively. The 840-bp genomic region, which is included between the transcription start sites of the early transcripts T_6 and T_5, is shown. The TSSs are indicated as thin black bent arrows. Transcripts T_{ind} and T_{lys} are represented as gray and black dashed lines, respectively. The previously mapped termination signals of T_{ind} and the putative ones of T_{lys} are indicated by gray and black bars. (B) The 5′-3′ nucleotide sequence of *f55*, as well as the immediately upstream and downstream regions, are shown. The TATA box and the BRE element are centered at −23 nt and −28 nt, respectively. Start and stop codons are in boldface and boxed, while predicted transcription termination signals are in boldface and underlined.

F55, producing two distinct DNA-protein complexes (A1 and A2). The band intensity was found to be proportional to the amounts of protein used in the reactions (from 3 to 16 μM), suggesting that there were two F55 binding sites in the probe (Fig. 6B).

In order to determine the F55 dissociation constant (K_d) of the DNA-protein complexes, the relative radioactive signals were quantified, and data were used to generate binding curves (Hill plots). We found that the K_d was 5.2 μM for the faster-migrating complex A1 ($K_{d\,A1}$) and 9.3 μM for the slower-migrating complex A2 ($K_{d\,A2}$) (Fig. 6B). A third signal A3 was very weak in intensity and only detectable at a high concentration of F55 (8 to 16 μM), which more likely represented unspecific DNA-protein aggregates. Consequently, we only studied the main A1 and A2 complexes and their equivalents further.

The A1 and A2 complexes exhibited different stability as confirmed by specific competition assays that abolished first the signal of A2 and then that of A1 (Fig. 6B). The shifts almost completely

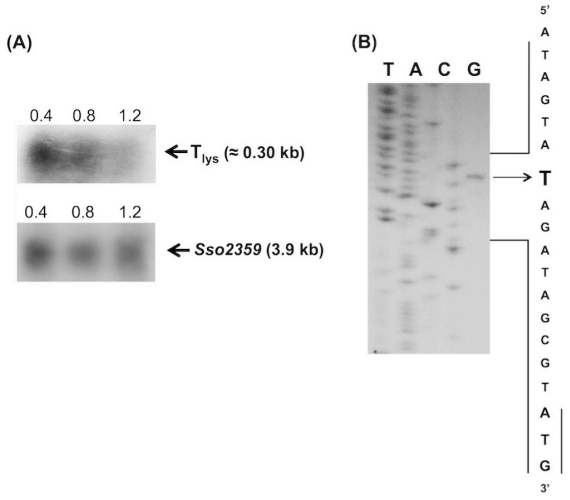


FIG 3 Transcription analysis of T_{lys}. (A) SSV1-InF1 total RNAs isolated at three different growth stages (OD_{600} = 0.4, 0.8, and 1.2) that highlight the variation of gene expression at the early exponential, late exponential, and stationary phases of growth, respectively, were analyzed by Northern blotting and primer extension experiments. The oligonucleotides used for the analysis of T_{lys} transcript is T_{lys}rv (see Material and Methods). The cDNA products were electrophoresed with the sequence ladder generated by the same primer on the noncoding strand of the *f55* gene. The mapped TSS is indicated by the arrow. Detection of the housekeeping gene *Sso2359* was performed as described above.

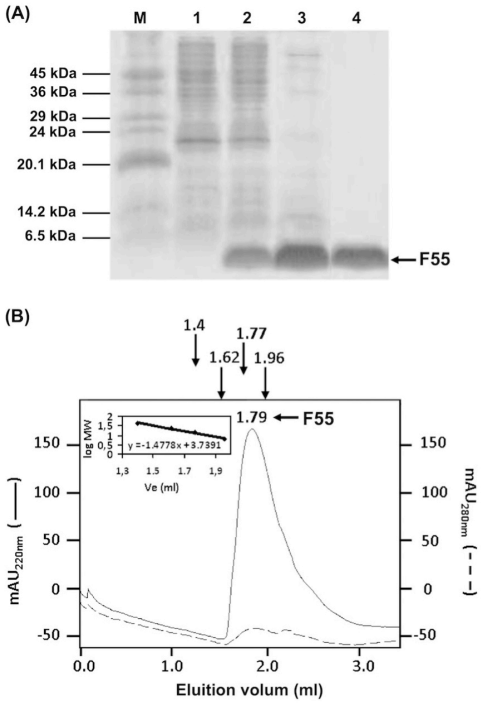


FIG 5 SDS-PAGE and quaternary structure analyses of the recombinant F55. (A) SDS-PAGE of protein extracts at each step of the protein purification. Lane M, molecular-mass markers; lane 1, crude extract from not induced cells; lane 2, crude extract from cells induced with 1 mM IPTG; lanes 3 and 4, samples after cationic exchange and gel filtration chromatography. (B) Elution profile from gel filtration chromatography on a Sephadex PC75 16/60 column. The elution volume of F55 (1.79 ml) corresponds to a molecular mass of 12.4 kDa. Arrows indicate the elution volumes of the protein standards in the relative calibration of the column.

disappeared at 500 molar excess of the cold DNA, as judged by quantification of the free probe signal (>90% of the initial amount).

The protein F55 binds to several operator sequences located in the promoters of the early transcripts. The high expression level of T_{lys} in the SSV1-InF1 lysogenic strain suggested that it could play a key role in controlling SSV1 lysogeny. To yield an insight into this hypothesis, the whole SSV1 genome was scanned to identify tandem-repeated sequences similar to those present in the T_{lys} promoter. This element was identified in three other promoters, i.e., of the early transcripts T_5 and T_6 and of T_{ind}, a transcript that is highly induced upon UV irradiation. Interestingly, all of these repeated sequences are found within a small region (~840 bp) of the SSV1 genome that is included between the promoter sequences of transcripts T_5 and T_6 (Fig. 6A).

Two copies of the tandem-repeated element were found in all of these promoters. T_5-A-TR and T_6-C-TR encompass the TSS of T_5 and T_6, respectively. Instead, T_5-B-TR and T_6-D-TR lie immediately upstream of the core promoters of these transcripts and

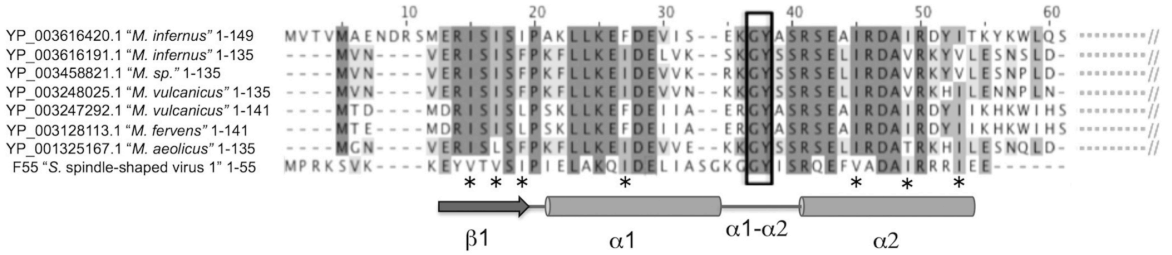


FIG 4 Multiple sequence alignment and secondary structure prediction of F55. The sequence of F55 aligns with the N termini of NikR proteins encoded by the genome of *Methanocaldococcus* spp. and *Methanococcus* spp. Residues conserved in all of the sequences analyzed are shaded in dark gray. The dipeptide Gly-Tyr (GY), typically located between helixes α1 and α2 in the RHH proteins, is boxed. The intensity of the gray color decreases proportional to the percentage of conservation. Asterisks indicate conserved hydrophobic, primarily branched-chain amino acids. The secondary structure prediction based on Jpred algorithm, is also schematized.

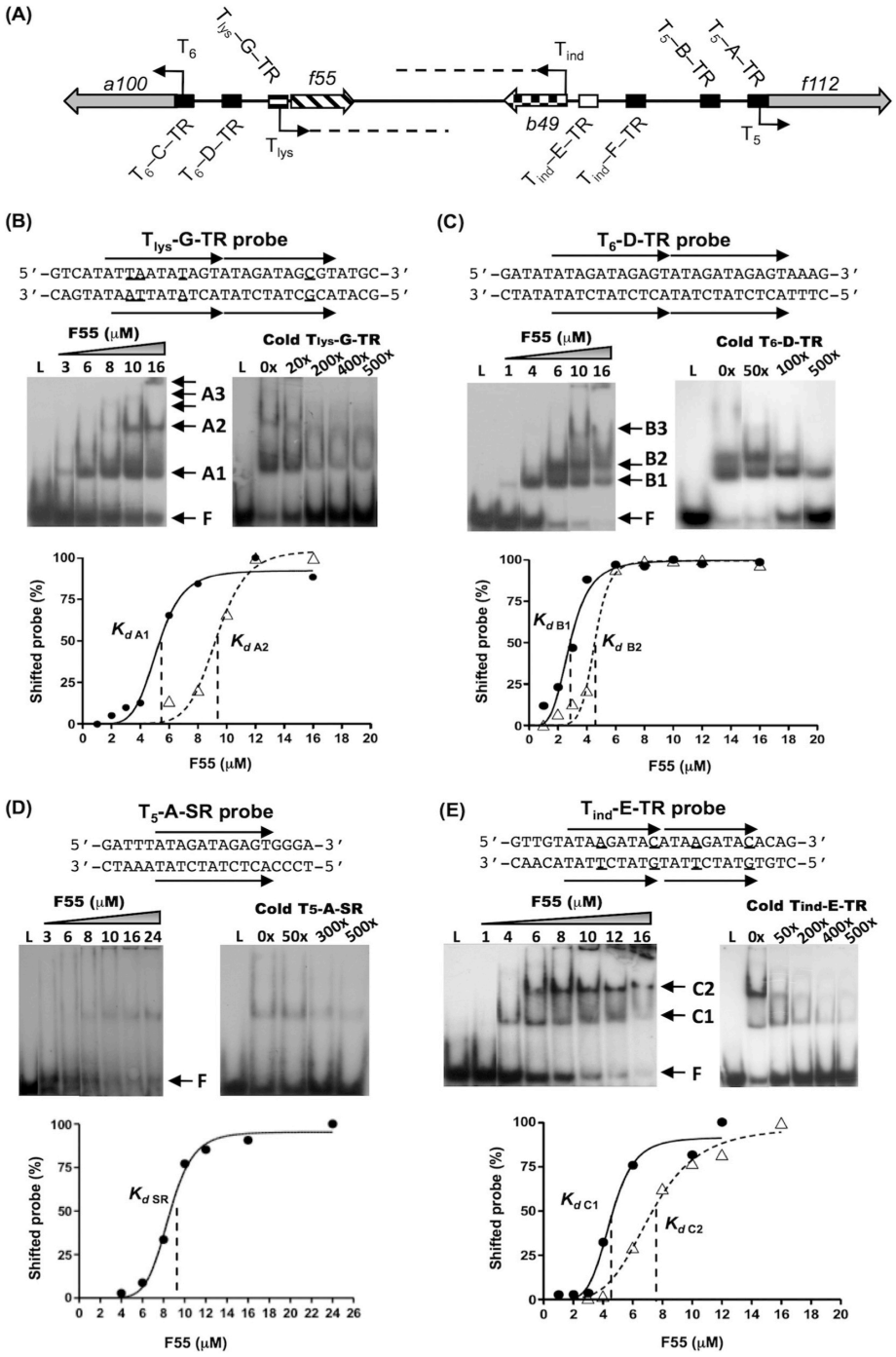
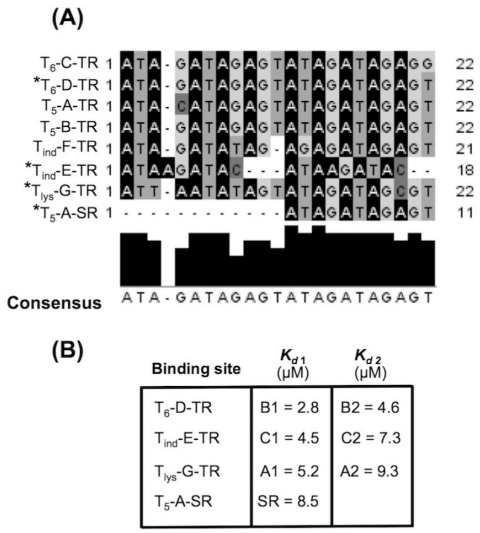


FIG 6 Analysis of the F55 interaction with operator sequences localized in the key regulative region of SSV1. (A) Graphic representation of the region included between the TSS of the early transcripts T_5 and T_6. ORFs are represented as straight arrows, and TSSs are indicated as thin black bent arrows. Tandem-repeated sequences (TR) are reported as black, white, gray, or striped boxes located upstream of the relative ORFs. (B to E) Band-shift assays results. For each experiment, the name and the sequence of the probe tested are indicated. Black arrows highlight the tandem-repeated sequences in which underlined letters indicate the mismatches compared to the consensus sequence. The two panels indicate the EMSAs performed with increasing concentrations of F55 (left) and with increasing concentrations of the specific cold probe (right), whose molar excess is reported on the top. Different DNA-protein complexes are indicated, and "F" means free probe. Binding curves relative to the formation of the DNA-protein complexes are reported. Densitometric data from EMSA obtained as described in Materials and Methods are plotted versus the concentration of F55.



(B)

Binding site	K_{d1} (μM)	K_{d2} (μM)
T_6-D-TR	B1 = 2.8	B2 = 4.6
T_{ind}-E-TR	C1 = 4.5	C2 = 7.3
T_{lys}-G-TR	A1 = 5.2	A2 = 9.3
T_5-A-SR	SR = 8.5	

FIG 7 Sequence alignment of the F55 binding sites and comparison of their dissociation constants. (A) F55 binding sites, the relative sequences, and the lengths are shown. Asterisks indicate the sequences tested in EMSAs. Nucleotides conserved in a specific position are highlighted in the same color, while substitutions are shown in a different color. The height of the black histograms is proportional to the percentage of the conservation of each nucleotide position; the consensus sequence is shown. (B) For the tested sequences, the K_d values relative to the faster-migrating (A1, B1, and C1) and slower-migrating (A2, B2, and C2) complexes are reported.

partially overlap their BRE elements (Fig. 6A). Despite the absence of canonical basal promoter elements (TATA box and BRE), the T_{ind} promoter region contains two copies of the tandem-repeated sequence that are similarly spaced as those of T_5 and T_6. The multiple alignment of these sequences (Fig. 7A) allowed the identification of a 22-bp consensus element (5′-ATAGATAGAGTATAG ATAGAGT-3′).

When T_6-D-TR was tested, two differently migrating complexes were identified. The complex B1 appeared to precede B2, indicating that the formation of B2 is a multistep process (Fig. 6C). The K_d values of the complexes ($K_{d\,B1}$ = 2.8 μM and $K_{d\,B2}$ = 4.6 μM), indicated a better affinity of F55 for T_6-D-TR than for T_{lys}-G-TR (Fig. 7B). The complexes B1 and B2 displayed different stability (B2 < B1) as inferred by displacement experiments with specific cold DNA (Fig. 6C). Comparing the results obtained with T_{lys}-G-TR (complexes A1 and A2), the faster-migrating signal B1, although decreased upon addition of cold specific DNA, still persists at 500 molar excess, confirming that the interaction between F55 and T_6-D-TR is stronger than that with T_{lys}-G-TR (Fig. 6B and C).

We also tested whether F55 could bind to a single consensus module (5′-ATAGATAGAGT-3′) of the tandem repeat (Fig. 7A). With this aim, the 20-bp probe T_5-A-SR (SR, i.e., single repeat) was labeled and used in band-shift assays as described above (Fig. 6D). Only one shifted signal was detected, indicating that one binding site of F55 is included in the single module. The construction of a binding curve revealed a higher dissociation constant ($K_{d\,SR}$ = 8.5 μM; Fig. 7B); therefore, the DNA-protein complex formed with a single module is less stable than the complexes formed with a dual module. Furthermore, specific displacement experiments indicate that F55 specifically binds to the sequence 5′-ATAGATAGAGT-3′ (Fig. 6D), which is present in slightly different variants in every promoter tested (Fig. 7A). According to these EMSA data, there is a correlation between the presence of mismatches in the tandem-repeated sequence (as in T_{lys}-G-TR) and the decreased affinity of F55 toward its targets.

To better define the target sequence of F55, T_{ind}-E-TR was used as a DNA substrate. A similar binding pattern with two shifted bands (C1 and C2) was observed (Fig. 6E). Further, $K_{d\,C1}$ and $K_{d\,C2}$ values were in the same range as those determined for T_{lys}-G-TR and T_6-D-TR (Fig. 7B). Interestingly, the formation of the complex C2 occurred at a F55 concentration that was lower than the $K_{d\,C1}$ (4.5 μM) (Fig. 6E, dashed curve), which contrasted to the results obtained with other probes. Indeed, for all of the other sequences tested, the appearance of the slowest-migrating complexes A2 and B2 (Fig. 6B and C, dashed curves) occurred as soon as the F55 concentration reached the K_d values of A1 and B1, respectively (Fig. 6B and C, solid curves). A possible explanation for this difference is that the two modules of the tandem-repeated sequence are closer to each other, thus influencing positively the interaction of F55 to the binding sites of T_{ind}-E-TR (Fig. 7A).

Displacement experiments performed with salmon testes DNA (data not shown) confirmed that F55 exhibited the highest binding specificity toward T_6-D-TR, which contains the perfect tandem-repeated sequence (Fig. 7A). However, under our experimental conditions F55 showed a not-pronounced difference in the apparent affinity to specific and nonspecific DNA, in accordance to what already reported for other RHH-containing transcription regulators, such as Lrp from *S. solfataricus* (37) and SvtR from virus SIRV1 (38).

Taken together, these findings show that F55 interacts with differential affinity to all of the identified repeated sequences in the early UV-inducible region of the SSV1 genome.

DISCUSSION

In this study, we have detected a transcriptional activity within the intergenic region ranging from the T_6 to the T_{ind} promoters. This novel transcript contains an ORF, which encodes a putative transcription factor (F55). This gene is highly expressed in the SSV1-InF1 strain, only in the absence of UV stimulus (lysogenic state), suggesting a key role for F55 in the SSV1 lysogeny. Intriguingly, the protein F55 binds specifically to a direct-repeated motif that recurs in its own promoter, as well as in the promoters of the early T_5 and T_6 (26, 27, 39), and of the UV-inducible T_{ind} transcripts.

While the UV induction of SSV1 has been characterized in the native and foreign hosts at physiological and transcriptional levels (17, 18, 26–29), the lysogenic growth of the virus has never been studied extensively. Nevertheless, our experimental evidences, such as the stringent control of the SSV1 copy number during the lysogenic growth, suggested the occurrence of a tight regulation of the SSV1 replication in the absence of UV stimulus. Therefore, we resolved to unravel the molecular components underpinning the maintenance of the lysogenic state.

The *f55* gene was first identified here. There are two reasons that could account for the delayed discovery of this gene: (i) when UV-inducible expression was investigated in the natural host, the expression of T_{lys} was inhibited (26), and (ii) in the microarray analysis by Fröls et al., no probe was designed for the above-mentioned intergenic region (27).

The primary structure of F55 shares a good degree of sequence

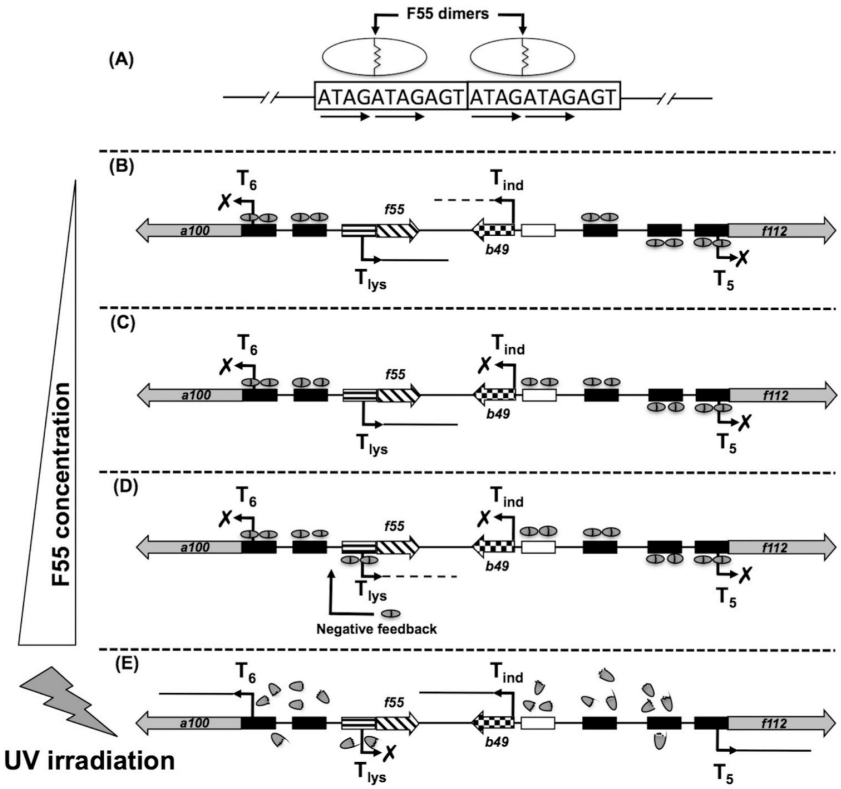


FIG 8 Model of the F55 interaction at its binding sites in the key regulative region. (A) The consensus tandem-repeated sequences is framed. Black arrows indicate the adjacent binding subsites (5′-ATAG-3′). Each F55 dimer (white ovals) binds to two adjacent subsites (5′-ATAG-3′) within a single repeat. (B to D) Progressive concentration-dependent saturation of the F55 binding sites as inferred from their K_d values. (B) At low concentrations, F55 binds to two tandem repeats of T_5 and T_6, as well as to one of the T_{ind} (black boxes). (C and D) As the F55 concentration increases, the second T_{ind} binding site (white box) and that of T_{lys} are saturated sequentially. ORFs and TSSs are indicated as large solid arrows and thin black bent arrows, respectively. The interaction of F55 at its binding sites causes complete transcriptional abrogation (black crosses) or downregulation (dashed lines starting from TSS). In this latter case (D), a negative-feedback regulation mechanism operates on the T_{lys} expression. (E) Upon UV irradiation, F55 is degraded and/or inactivated and, in turn, the transcription of the early induced T_{ind}, T_5, and T_6 transcripts is unlocked.

identity with several transcription repressors such as those belonging to NikR or CopG families. These regulators form a dimeric DNA-binding domain constituted by a pair of antiparallel β-strands framed within an α-helical scaffold (34, 35). Accordingly, our secondary structure prediction indicated that F55 also folds into the RHH motif and size exclusion chromatography demonstrated that it exists as dimer in solution.

Some examples of archaeal RHH-containing proteins have been reported and are: (i) ORF56, encoded by the *S. islandicus* pRN1 plasmid (40), (ii) the regulator SvtR encoded by the rudivirus SIRV1 (38), (iii) the protein E73 encoded by the virus SSV-RH (16), and (iv) the recently described AvtR encoded by the lipothrixvirus AFV6 (41). Typically, multiple RHH dimers bind to an array of regularly spaced inverted or tandem repeats in the promoters of pertinent gene(s), which include their own genes, to exert transcriptional repression (42).

Here we show that the dimeric F55 binds specifically to an 11-nt sequence (5′-ATAGATAGAGT-3′), which is the minimal binding site, leading to the formation of a stable DNA-protein complex. This sequence motif contains in turn a tandem-repeated element which fits well with the 2-fold symmetry of the dimeric $(RHH)_2$ domain that requires two adjacent subsites (5′-ATAG-3′ in this case) to bind the DNA (Fig. 8A).

In the hypothesis that F55 were a repressor involved in the maintenance of the lysogenic state by downregulating the transcription of the early induced transcripts (T_5, T_6, and T_{ind}), as well as its own expression, we tested the ability of F55 to bind to the relative promoter sequences.

While the promoters of T_5 and T_6 and of the UV-inducible T_{ind} transcripts bear four F55 binding sites arranged as two distinct 22-nt tandem-repeated sequences, the promoter of T_{lys} carries only one element overlapping its own TSS. Since the F55 binding sites present in promoters of T_5 (T_5-A-TR and T_5-B-TR), T_6 (T_6-C-TR and T_6-D-TR) and of T_{ind} (T_{ind}-F-TR) are overall identical to one another (Fig. 8B), we resolved to analyze only T_6-D-TR. The recombinant F55 interacts more strongly to the above-mentioned binding sites if compared to that of T_{ind} (T_{ind}-E-TR) and of its own promoter (T_{lys}-G-TR). This is in agreement with the fact that the former binding sites are very similar to the consensus sequence; instead, T_{ind}-E-TR and T_{lys}-G-TR differ significantly.

The substantial divergence of T_{ind}-E-TR from the consensus provides a molecular basis for the apparent cooperativity in the binding observed specifically after quantification of the relative EMSA signals. This DNA-binding mode is in accordance with the fact that T_{ind} is the first transcript to be produced upon the onset of the UV response and therefore transcriptional regulation at the T_{ind} promoter needs to be finely tuned and quickly reverted.

As in *Bacteria*, the placement of the binding sites of transcription repressors relative to promoter elements is also the primary determinant affecting transcription initiation in *Archaea* (43). The location of the F55 binding sites overlapping the TSS and the BRE element suggests a negative regulatory role of these *cis*-acting elements recognizable by transcription repressor(s) (44). Indeed, analogous to previously characterized archaeal negative regulators such as MDR1 from *Archaeoglobus fulgidus* (45) or LrpA from *Pyrococcus furiosus* (46), F55 may impinge on the activity of the transcription machinery through binding to operator sequences that overlap the transcription start site, thus abrogating RNAp (RNA polymerase) recruitment to the promoter and preventing initial steps of RNA chain elongation. Alternatively, by interacting to operator sequences that overlap the BRE element, F55 might convey its repressive effect on transcription initiation by hindering an earlier step, i.e., the formation of the ternary TBP-TFB-DNA PIC (preinitiation complex). This kind of transcriptional repression mode has been already described for the *Sulfolobus solfataricus* Lrs-14 protein (47) and for the *Thermococcus litoralis* TrmB transcription factor (48).

In vivo analysis indicates that the ORF *f55* is strongly downregulated during the SSV1 life cycle and reaches its lowest expression level in the late-stationary phase, thus suggesting that the protein F55 is a transcription repressor that modulates its own expression. This strategy allows the cell to avoid overproducing the repressor protein and to keep its functional concentration.

The newly identified *f55* gene is located in an SSV1 genomic region which resembles that of the early induced UV genes of bacteriophage lambda. In this case the intergenic region, located between the early genes, encodes the *c*I repressor protein. This regulator simultaneously represses transcription of the early induced genes by binding to specific operator regions, thus allowing the maintenance of the lysogenic state to be governed by *c*I alone.

The lysogenic growth of lambda requires that this repressor binds to all three operators at the left and right promoters, and its oligomerization on the target sites shut off the promoters completely (20). We hypothesize that F55 controls the SSV1 lysogeny in a similar fashion. Although our data do not indicate whether F55 exerts its physiological function through binding to one or both tandem-repeated sequences within the same promoter, oligomerization most likely starts from the most affine primary sites and subsequently extend to the entire region in a concentration-dependent manner. Therefore, according to its differential affinity toward the binding sites tested (Fig. 7B), F55 would bind first to the sequences in the promoters of T_5, T_6, and T_{ind} transcripts (Fig. 8B, black boxes) and then extend to the less affine sequence of T_{ind} (Fig. 8C, white box) and finally to its own promoter (Fig. 8D, striped box), possibly repressing the expression of the corresponding transcripts, sequentially. Saturation of all of the binding sites may serve to enhance transcription repression.

By analogy to the lambda phage (20), it is tempting to speculate that F55 is degraded and/or inactivated upon UV irradiation in a fashion similar to *c*I, thus unblocking the transcriptional circuit of the early genes (Fig. 8E).

In contrast to a lambda lysogen, in which episomic DNA is completely absent, both episomic and integrated forms of SSV1 exist in SSV1 lysogenic cells (17). This indicates that SSV1 has to express a minimal set of proteins responsible for a low level of DNA replication and/or for the maintenance of the lysogenic state. The absence of F55 binding sites in the promoters of the T_x, T_3, and T_9 transcripts, which are expressed together with T_{lys} during the lysogenic growth (data not shown), further support the hypothesis that F55 functions as a regulator only toward the early UV-inducible transcripts.

ACKNOWLEDGMENTS

We thank the *Sulfolobus* Gene Chip Consortium coordinated by John van der Oost for constructing the microarrays used in this study and Luciano Pirone for performing the light-scattering analysis of F55. We are grateful to Gabriella Fiorentino for helpful scientific discussions.

This study was supported by Danish grant 11-106683 from Danish Independent Research Council–Technology and Production Sciences and from Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica (Progetti di Ricerca di Interesse Nazionale E61J10000020001).

REFERENCES

1. **Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W.** 1984. SAV1, a temperate UV-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J. **3**:2165–2168.
2. **Wiedenheft B, Stedman KM, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M.** 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J. Virol. **78**:1954–1961.
3. **Redder P, Peng X, Brugger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D.** 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ. Microbiol. **11**:2849–2862.
4. **Stedman KM, Schleper C, Rumpf E, Zillig W.** 1999. Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. Genetics **152**:1397–1405.
5. **Cannio R, Contursi P, Rossi M, Bartolucci S.** 1998. An autonomously replicating transforming vector for *Sulfolobus solfataricus*. J. Bacteriol. **180**:3237–3240.
6. **Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, Bartolucci S.** 2003. Development of a genetic system for hyperthermophilic *Archaea*: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. FEMS Microbiol. Lett. **218**:115–120.
7. **Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C.** 2003. A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. Mol. Microbiol. **48**:1241–1252.
8. **Reiter WD, Palm P, Henschen A, Lottspeich F, Zillig W, Grampp B.** 1987. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. Mol. Genet. Genomics **206**:144–153.
9. **Prangishvili D, Garrett RA, Koonin EV.** 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res. **117**:52–67.
10. **Kraft P, Kummel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM.** 2004. Structure of D-63 from *Sulfolobus* spindle shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J. Virol. **78**:7438–7442.
11. **Helmer-Citterich M, Anceschi MM, Banner DW, Cesareni G.** 1988. Control of ColE1 replication: low affinity specific binding of Rop (Rom) to RNAI and RNAII. EMBO J. **7**:557–566.
12. **Kraft P, Oeckinghaus A, Kummel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM.** 2004. Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA-binding protein. J. Virol. **78**:11544–11550.

13. **Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzerd E, Young MJ, Bothner B, Lawrence CM.** 2008. Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology **376**:270–278.
14. **Wu RY, Zhang RG, Zagnitko O, Dementieva I, Maltzev N, Watson JD, Laskowski R, Gornicki P, Joachimiak A.** 2003. Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. J. Biol. Chem. **278**: 20240–20244.
15. **Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF.** 2001. The crystal structure of MarR a regulator of multiple antibiotic resistance, at 2.3-Å resolution. Nat. Struct. Biol. **8**:710–714.
16. **Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakić M, Young MJ, Lawrence CM, Copié V.** 2012. Structural studies of E73 from a hyperthermophilic archaeal virus identify the "RH3" domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. Biochemistry **51**: 2899–2910.
17. **Schleper C, Kubo K, Zillig W.** 1992. The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc. Natl. Acad. Sci. U. S. A. **89**:7645–7649.
18. **Ceballos RM, Marceau CD, Marceau JO, Morris S, Clore AJ, Stedman KM.** 2012. Differential virus host-ranges of the *Fuselloviridae* of hyperthermophilic *Archaea*: implications for evolution in extreme environments. Front. Microbiol. **3**:295 doi:10.3389/fmicb.2012.00295.
19. **Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I.** 1980. The *Sulfolobus*-"*Caldariella*" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch. Microbiol. **125**:259–269.
20. **Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S.** 2005. Switches in bacteriophage lambda development. Annu. Rev. Genet. **39**: 409–429.
21. **Contursi P, Jensen J, Aucelli T, Rossi M, Bartolucci S, She Q.** 2006. Characterization of the *Sulfolobus* host-SSV2 virus interaction. Extremophiles **10**:615–627.
22. **Contursi P, Cannio R, Prato S, She Q, Rossi M, Bartolucci S.** 2007. Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. J. Bacteriol. **17**:6339–6350.
23. **Contursi P, Cannio R, She Q.** 2010. Transcription termination in the plasmid/virus hybrid pSSVx from *Sulfolobus islandicus*. Extremophiles **14**: 453–463.
24. **She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weihere CCY, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PMK, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J.** 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc. Natl. Acad. Sci. U. S. A. **98**:7835–7840.
25. **She Q, Brugger K, Chen L.** 2002. Archaeal integrative elements and their impact on evolution. Res. Microbiol. **153**:325–332.
26. **Reiter WD, Palm P, Yeats S, Zillig W.** 1987. Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. **209**:270–275.
27. **Fröls S, Gordon PM, Panlilioc MA, Schleper C, Sensen CW.** 2007. Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology **365**:48–59.
28. **Reiter WD, Palm P, Zillig W.** 1988. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res. **16**:1–19.
29. **Reiter WD, Palm P, Zillig W.** 1988. Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. Nucleic Acids Res. **16**:2445–2459.
30. **Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, Garrett RA.** 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol. Microbiol. **79**:35–49.
31. **Lee C, Kim J, Shin SG, Hwang S.** 2006. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. J. Biotechnol. **123**:273–280.
32. **Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ.** 2009. Jalview version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics **25**:1189–1191.

32a. **Prangishvili D, Forterre P, Garrett RA.** 2006. Viruses of the *Archaea*: a unifying view. Nat. Rev. Microb. **4**:837–848.

33. **Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W.** 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology **185**:242–250.
34. **Chivers PT, Sauer RT.** 2000. Regulation of high-affinity nickel uptake in bacteria. J. Biol. Chem. **275**:19735–19741.
35. **Schreiter ER, Drennan CL.** 2007. Ribbon-helix-helix transcription factors: variations on a theme. Nat. Rev. Microbiol. **5**:710–720.
36. **Contursi P, D'Ambrosio K, Pirone L, Pedone E, Aucelli T, She Q, De Simone G, Bartolucci S.** 2011. C68 from the *Sulfolobus islandicus* plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. Biochem. J. **435**:157–166.
37. **Stacey SN, Dahlquist FW, Reich NO.** 2007. The role of high-affinity nonspecific DNA binding by Lrp in transcriptional regulation and DNA organization. J. Mol. Biol. **369**:1307–1317.
38. **Guillière F, Peixeiro N, Kessler A, Raynal B, Desnoues N, Keller J, Delepierre M, Prangishvili D, Sezonov G, Guijarro JI.** 2009. Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. J. Biol. Chem. **284**:22222–22237.
39. **Qureshi SA.** 2007. Protein-DNA interactions at the *Sulfolobus* spindle-shaped virus-1 (SSV1) T5 and T6 gene promoters. Can. J. Microbiol. **53**: 1076–1083.
40. **Lipps G, Stegert M, Krauss G.** 2001. Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRN1, a putative archaeal plasmid copy control protein. Nucleic. Acids Res. **29**:904–913.
41. **Peixeiro N, Keller J, Collinet B, Leulliot N, Campanacci V, Cortez D, Cambillau C, Nitta KR, Vincentelli R, Forterre P, Prangishvili D, Sezonov G, van Tilbeurgh H.** 2013. Structure and function of AvtR, a novel transcriptional regulator from a hyperthermophilic archaeal lipothrixvirus. J. Virol. **87**:124–136.
42. **Zampini M, Hayes F.** 2012. Combinatorial targeting of ribbon–helix–helix artificial transcription factors to chimeric recognition sites. Nucleic Acids Res. **40**:6673–6682.
43. **Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G.** 2009. Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in *Archaea* by MarR family proteins. J. Mol. Biol. **388**: 559–569.
44. **Fiorentino G, Cannio R, Rossi M, Bartolucci S.** 2003. Transcriptional regulation of the gene encoding an alcohol dehydrogenase in the archaeon *Sulfolobus solfataricus* involves multiple factors and control elements. J. Bacteriol. **185**:3926–3934.
45. **Bell SD, Cairns SS, Robson RL, Jackson SP.** 1999. Transcriptional regulation of an archaeal operon in vivo and in vitro. Mol. Cell **4**:971–982.
46. **Brinkman AB, Dahlke I, Tuininga JE, Lammers T, Dumay V, de Heus E, Lebbink JH, Thomm M, de Vos WM, van Der Oost J.** 2000. An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. J. Biol. Chem. **275**:38160–38169.
47. **Bell SD, Jackson SP.** 2000. Mechanism of autoregulation by an archaeal transcriptional repressor. J. Biol. Chem. **275**:31624–31629.
48. **Lee SJ, Engelmann A, Horlacher R, Qu Q, Vierke G, Hebbeln C, Thomm M, Boos W.** 2003. TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. J. Biol. Chem. **278**:983–990.

2.3 A standardized protocol for the UV induction of *Sulfolobus* spindle-shaped virus 1

Extremophiles
DOI 10.1007/s00792-014-0717-y

METHOD PAPER

A standardized protocol for the UV induction of *Sulfolobus* spindle-shaped virus 1

Salvatore Fusco · Martina Aulitto · Simonetta Bartolucci · Patrizia Contursi

Received: 29 September 2014 / Accepted: 26 November 2014

Abstract The *Fuselloviridae* prototype member *Sulfolobus* spindle-shaped virus 1 is a model of UV-inducible viruses infecting *Crenarchaeota*. Previous works on SSV1 UV induction were bases on empirically determined parameters that have not yet been standardized. Thus, in many peer reviewed literature, it is not clear how the fluence and irradiance have been determined. Here, we describe a protocol for the UV induction of SSV1 replication, which is based on the combination of the following instrumentally monitored parameters: (1) the fluence; (2) the irradiance; (3) the exposure time, and (4) the exposure distance. With the aim of finding a good balance between the viral replication induction and the host cells viability, UV-irradiated cultures were monitored for their ability to recover in the aftermath of the UV exposure. This UV irradiation procedure has been set up using the well-characterized *Sulfolobus solfataricus* P2 strain as model system to study host–virus interaction.

Keywords *Sulfolobus* spindle-shaped virus · UV induction · Irradiation protocol · *Fuselloviridae* · Viral titre · Plaque assay · Fluence · Irradiance

Communicated by L. Huang.

Electronic supplementary material The online version of this article (doi:10.1007/s00792-014-0717-y) contains supplementary material, which is available to authorized users.

S. Fusco · M. Aulitto · S. Bartolucci · P. Contursi (✉)
Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, 80126 Naples, Italy
e-mail: contursi@unina.it

Introduction

Sulfolobus spindle-shaped virus 1 (SSV1), isolated from the native host *Sulfolobus shibatae* in Beppu (Japan), is the prototype and the best characterized member of the *Fuselloviridae* family (Prangishvili 2013; Martin et al. 1984). SSV1 can propagate only in few hosts (Ceballos et al. 2012), among which a strain of *S. solfataricus* isolated from the solfataric field of Pisciarelli near Naples (Italy), turned out to be suitable for carrying out genetic (Stedman et al. 1999; Clore and Stedman 2007; Iverson and Stedman 2012), biochemical (Kraft et al. 2004a, b; Menon et al. 2008; Zhan et al. 2012; Eilers et al. 2012) and physiological studies (Reiter et al. 1987; Schleper et al. 1992; Zillig et al. 1980; Fröls et al. 2007a; Fusco et al. 2013).

The genome of SSV1 is a double-stranded DNA molecule of 15 Kbp, which has been found both as integrated (provirus) and as episomal form into the host cells (Schleper et al. 1992; Yeats et al. 1982). Its complete sequence has been determined (Palm et al. 1991), and as for other fuselloviruses (Stedman et al. 2003; Wiedenheft et al. 2004; Redder et al. 2009; Contursi et al. 2007, 2010, 2014a), it encodes for a number of quasi-orphan proteins, which do not have detectable homologues in the databases other than in related hyperthermophilic viral genomes (Contursi et al. 2013). This has led to the necessity of performing structural and functional analyses to unravel their functions. For instance, the structure of several SSV1 transcription factors (TFs) has been solved revealing that, despite the lack of homology, most of these viral TFs are bacterial like (Kraft et al. 2004a, b; Menon et al. 2008; Zhan et al. 2012; Eilers et al. 2012; Contursi et al. 2013; 2011; 2014b).

Insights about the role of some SSV1 proteins have been derived from genetic analyses that revealed some of the

Published online: 06 December 2014

essential SSV1 genes (Stedman et al. 1999; Iverson and Stedman 2012). Furthermore, its genome has been used as template for the construction of replicative and expression vectors, which have been employed for heterologous gene expression as well as for studying viral ORFs essentiality (Contursi et al. 2014a; Stedman et al. 1999; Iverson and Stedman 2012; Jonuscheit et al. 2003; Albers et al. 2006; Cannio et al. 1998; Cannio et al. 2001).

Pioneering studies on SSV1 helped shedding light on how gene expression is regulated in *Archaea*. In particular, the elucidation of the SSV1 transcriptional map as well as the identification of all the transcriptional start sites (TSSs) (Reiter et al. 1987; Fröls et al. 2007a; Fusco et al. 2013), led to the discovery of two conserved sequence motifs that resemble those of the eukaryotic basal gene promoters recognized by the RNA polymerase II (Reiter et al. 1988). Therefore, the bacterial-like transcription regulators encoded by *Archaea* operate in a eukaryal-like transcriptional context (Contursi et al. 2013).

Upon infection, one copy of the viral genome site specifically integrates into the host chromosome at an arginyl-tRNA gene (Schleper et al. 1992), whereas the episomal form (~5 copies per cell) is maintained in host cells in three isoforms, i.e. (1) as positively or negatively supercoiled and (2) as relaxed double-stranded DNA (Snyder et al. 2003). Intriguingly, so far SSV1 is the only member of the *Fuselloviridae* family that shows an UV-inducible life cycle. Upon UV light exposure a well-characterized gene expression pattern is triggered and involves the expression of a short UV-inducible transcript, namely T_{ind}, followed by: (1) the time-coordinated expression of all the other viral transcripts, (2) the induction of the SSV1 genome replication (Fröls et al. 2007a) and (3) a steep increase of the viral titre. Only recently efforts have been made to get insights into the molecular switch from the lysogenic state to the replication induction (Fusco et al. 2013). Nevertheless, mechanisms underpinning these processes are still murky.

So far, in peer reviewed published literature, it is not clear how the fluence (or UV dose) and the irradiance have been measured. Indeed, the fluence (J m^{-2}) administered to the cells has been only empirically determined and no attention has been paid to monitoring the irradiance (J m^{-2} s^{-1}) (Martin et al. 1984; Reiter et al. 1987; Schleper et al. 1992; Fröls et al. 2007a). Furthermore, the negative effect on the host viability, that the UV treatment implies, has been underestimated. Since an essential parameter such as the irradiance has not been taken into account elsewhere (Martin et al. 1984; Reiter et al. 1987; Schleper et al. 1992; Fröls et al. 2007a) and the nomenclature may be misleading, important definitions need to be discussed. The irradiance is a proper term used when a surface is irradiated by UV light coming from all directions above the aforementioned surface. Indeed, the irradiance is the total radiant power incident from all upward directions on an infinitesimal element of surface having area dA and containing the point under consideration divided by dA, as defined by Bolton (Bolton and Linden 2003). When the irradiance is constant (as herein described), the fluence is derived multiplying the irradiance by the exposure time (in seconds). The term UV dose, so far used to refer to the fluence, is confusing since the word "dose" describes a total energy adsorbed by a surface (e.g. the skin). In UV irradiation of microbial suspensions, the major amount of the incident UV light crosses the sample with only a small percentage being adsorbed by the cells. For this reason, fluence is a much more appropriated term since it is related to the incident UV energy, rather than to the adsorbed fraction (Bolton and Linden 2003).

Herein, we describe a protocol that has been developed with the purpose of: (1) standardizing all the parameters needed for performing an UV induction experiment on SSV1 lysogens and (2) finding a good balance between the viral induction and the host viability. Indeed, we show that by tuning fluence and irradiance, cells viability can be improved and, in turn, the viral induction reaches highest values determined so far.

Materials and methods

Growth conditions and UV irradiation

Cells of the *S. solfataricus* P2 lysogenic strain (SSV1-P2) were revitalized by depositing few microliters of culture onto the soft layer of an SCVYU-Gelrite plate and incubating at 75 °C for 3–5 days, as described elsewhere (Contursi et al. 2006). Subsequently, local growth areas (spots) were inoculated into 50 ml of SCVY medium, i.e. a glycine-buffered Brock's basal salt solution, supplemented with 0.2 % sucrose (wt/vol), 0.2 % casamino acids (wt/vol), 1 × vitamins (Wolin et al. 1963) and 0.005 % yeast extract (wt/vol); the pH was adjusted to 3.5 with concentrated H_2SO_4. Cells cultivation was conducted in a 250-ml Erlenmeyer flask with a long neck, at 75 °C with a shaking rate of 180 rpm using a MaxQ™ 4000 Benchtop Orbital Shaker (Thermo Scientific). The cell growth was spectrophotometrically monitored at 600 nm (OD_{600}) by means of a Varian Cary® 50 Bio UV/Visible Spectrophotometer (McKinley Scientific). Once the culture reached the logarithmic phase of growth (0.4–0.6 OD_{600}), it was diluted to a value of 0.05 OD_{600} in 50 ml of fresh SCVY medium and let to grow up to 0.8 OD_{600}.

Before performing UV irradiations, the SSV1-P2 culture was diluted in 400 ml of SCVY medium to 0.08 OD_{600} into a 1-L Erlenmeyer flask and let to grow up to the mid-logarithmic growth phase (0.5 OD_{600}). Aliquots of this culture were then UV-irradiated or mock treated. In detail, 40 ml of

culture was transferred into a 150 × 25-mm Petri plate (BD Falcon™) and UV irradiation was carried out at room temperature in a dark room under red light, by carefully hand shaking the plate under a germicidal lamp G15T8 (254 nm, 15 W, Sankyo Denki). Differently from previous reports (Martin et al. 1984; Reiter et al. 1987; Schleper et al. 1992; Fröls et al. 2007a), the fluence and the irradiance were instrumentally measured by means of a Quantum-photoradiometer HD9021, equipped with an LP9021 UVC probe (Delta Ohm). The fluence was of 30, 45 and 60 J m^{-2} (with an irradiance of 1.0 J m^{-2} s^{-1}) or of 45, 60 and 75 J m^{-2} (with an irradiance of 0.5 J m^{-2} s^{-1}) (Fig. 1). The irradiance was tuned by changing the exposure distance, which was of 70 cm (for an irradiance of 0.5 J m^{-2} s^{-1}) or of 50 cm (for an irradiance of 1.0 J m^{-2} s^{-1}), whereas the desired fluence was achieved by changing the exposure time to the UV light source. As control, a mock-treated sample was subjected to the same procedure except for the UV light exposure. Treated samples were separately collected in 250-ml Erlenmeyer flasks, which were wrapped with aluminium foil to protect the culture from further light exposure, and incubated at 75 °C with a shaking rate of 180 rpm.

To check cells viability immediately after the treatment, serial dilutions of the cultures were plated on SCVYU-Gelrite and incubated at 75 °C for 7–10 days. Colonies were counted (100–300 cells per plate) and a survival percentage was calculated. Moreover, cell growth was spectrophotometrically monitored throughout the post-treatment incubation and samples were taken after 8 and 24 h, because a peak in the amount of SSV1 DNA and in the viral titre was expected, respectively (our unpublished data). Cellular pellets as well as cell-free supernatants were obtained through centrifugation at 3,000×*g* for 15 min using the Centrifuge 5810R (Eppendorf). The procedure was carried out in triplicate and average (Avg) as well as standard deviation (SD) were calculated for the data reported below.

Quantitative plaque assay

SSV1 viral titre was determined for cell-free supernatants by quantitative plaque assays using the uninfected *S. solfataricus* strain P2, as lawn. This strain was first revitalized on SCVY-plate and then transferred into 50 ml of SCVYU medium, as described above. Cell density was monitored spectrophotometrically at 600 nm until the late logarithmic phase of growth (0.6 OD_{600}).

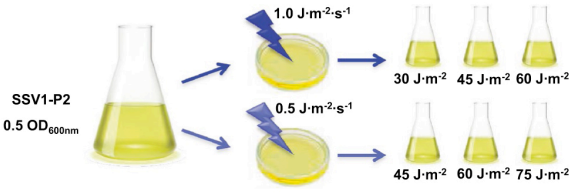


Fig. 1 Schematic illustration of the UV irradiation procedure. Cells were grown exponentially until 0.5 OD_{600} before being irradiated with an irradiance of 0.5 or 1.0 J m^{-2} s^{-1}

Lower layers of SCVY-Gelrite were prepared by pouring, in 100 × 15 mm plastic plates (Falcon), 30 ml of 1 × SCVY mixed with Gelrite® (Sigma Aldrich) at the final concentration of 0.8 % (w/v). Subsequently, 100 µl of serial dilutions (from 10^{-4} to 10^{-8}) of cell-free supernatants (containing SSV1 viral particles) were added to a mix composed of: (1) 1 × SCVY medium, (2) Gelrite® at the final concentration of 0.4 % (w/v) and (3) 0.5 ml of the 0.6 OD_{600} *S. solfataricus* P2 culture (about 0.5 × 10^8 cells). Each mix (the upper layer) was poured onto the lower layer of a pre-warmed SCVY-Gelrite plate. After a short incubation at room temperature to allow gelification of the upper layers, plates were transferred to 75 °C for 5–7 days. Growth inhibition areas (turbid halos), which appeared onto the upper layer as consequence of local growth retardation, were counted (up to 100 plaques per plate) and the viral titre (PFU/ml) was calculated considering the dilution factor.

Although quantitative plaque assay of SSV1 viral particles is notoriously challenging, we have noticed that a critical point for obtaining clearer halos depends on the physiological state as well as on the number of the cells used as lawn. Indeed, when about 0.5 × 10^8 cells of a not-freshly diluted culture are used, clear halos appear on the plate surface upon infection (Supplementary material, S1). Conversely, if the culture is freshly diluted with pre-warmed medium before plating, halos appear turbid and difficult to be counted.

Semi-quantitative PCR analysis

SSV1-P2 pellets, collected 8 h post-irradiation, were treated for total DNA extraction using the DNeasy tissue kit (Qiagen), following the manufacturer's instructions. The concentration of the DNA samples was spectrophotometrically measured by means of a Nanodrop 2000 Spectrophotometer (Thermo Scientific). To detect variations of the viral DNA content, total DNA samples from mock-treated and UV-irradiated cells were analysed by semi-quantitative PCR assays. With this aim, two primer couples were designed (Fusco et al. 2013) using Primer3 software (available at the website: http://bioinfo.ut.ee/primer3-0.4.0/), to amplify: (1) a 155-bp region of the SSV1 single-copy gene *vp*2 and (2) a 108-bp region of the host single-copy gene *orc*1 (Table 1). A PCR master mix was prepared as follows: 1 × Taq buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP mix (Thermo Scientific), 0.6 µM *orc*1-fw, 0.6 µM *orc*1-rv, 0.6 µM *vp*2-fw, 0.6 µM *vp*2-rv and 0.1U/µl of Taq DNA Polymerase (Thermo Scientific). Aliquots of the master mix (60 µl each) were dispensed into 200-µl tubes (Eppendorf) and 100 ng of total DNA from

Table 1 Sequences of oligonucleotides used for the semi-quantitative PCR assays

Name	Sequence (5′-3′)	Length (nt)
orc1-fw	TATAAATTGTTATAGACATAGAACGCTGTA	30
orc1-rv	TTAAATACTTCTTGTGCCGATAGTCC	26
vp2-fw	GGAGGGTACATCGCTACCTTATGA	24
vp2-rv	CAGTAGGGCTGACAGTAAACTACG	24

mock-treated or UV-irradiated cells were added. Each aliquot was then split into three sub-aliquots (20 µl each), to collect the tubes at the 20th, 25th and 30th cycle of amplification. The thermal cycling protocol was carried out into a Mastercycler Personal (Eppendorf®) as follows: an initial denaturation step of 5 min at 95 °C, followed by 30 cycles of 40 s at 95 °C, 40 s at 62 °C, and 1 min at 72 °C. After collecting the samples at the 30th cycle, a final step at 72 °C has been carried out for 10 min. PCR products were run on a 2 % agarose gel in 1 × TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) for 1 h. Pictures were taken by means of a Gel Doc XR System (Biorad) and DNA bands quantified using the Quantity One Software (Biorad).

Results

The UV irradiation has a dose-dependent effect on the host cell viability

The UV irradiation not only induces the viral replication in SSV1-lysogens, but affects metabolism and survival of the host as well. In fact, major changes in growth rate and gene expression have been observed in the aftermath of UV irradiation of *S. solfataricus* cells.

Nevertheless, discrepant data have been produced about the UV-dependent gene regulation in this crenarchaeon, probably because the equipment and the procedure used, have not been standardized (Fröls et al. 2007b, 2009; Götz et al. 2007; Salerno et al. 2003).

Likewise most of the bacteriophages and viruses known, the SSV1 development seems to be sensitive to the physiological state of the host (Prangishvili 2013; Contursi et al. 2006; Bondy-Denomy and Davidson 2014). Therefore, a suitable protocol for SSV1 induction requires taking into account the host response in regard of cell viability and ability to recover upon UV exposure.

With the aim of setting up the best conditions for the induction of the SSV1 replication, cells were UV-irradiated by combining a set of different parameters: (1) the fluence ranging from 30 to 75 J m^{-2}, (2) the irradiance of 0.5 or 1.0 J m^{-2} s^{-1}; (3) the exposure times ranging from 30 to 150 s, and (4) the exposure distance of 50 or 70 cm.

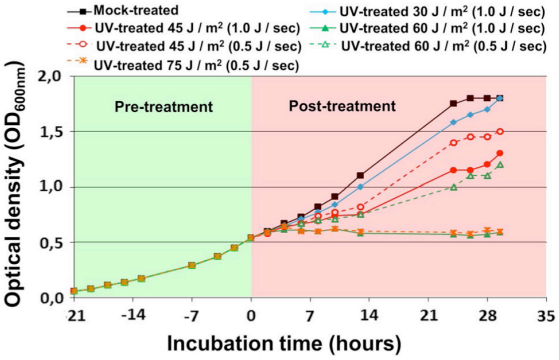


Fig. 2 Growth curves of SSV1-P2 pre- and post-UV treatment. The OD_{600nm} values were measured over a time window of ~55 h. Cells were grown exponentially until 0.5 OD_{600nm} value (21st h of incubation) before being mock or UV treated (*green* area of the graph). Afterwards, the samples were split into seven flasks, incubated back to 75 °C and further monitored (*red* area of the graph). The growth retardation is related to the fluence and irradiance administered

As shown in Fig. 2, cell growth is slowed down by the treatment with a dose-dependent trend. The highest UV fluency levels, i.e. 60 J m^{-2} and 75 J m^{-2}, resulted in heavy growth retardation, probably as consequence of an impaired activation of the DNA lesions repairing system(s). Conversely, cell growth was only partially delayed when the fluence and irradiance were progressively reduced down (Fig. 2). Furthermore, cells viability determined by plating aliquots of the cultures, revealed that the survival percentage upon UV irradiation gradually increased by reducing both the fluence and irradiance (Table 2). To compensate for effects related to temperature changes, control cultures were subjected to the same procedure except for the UV irradiation.

Interestingly, our data indicate that the fluence (J m^{-2}) is not the only parameter affecting the cell survival (Table 2) and their ability to recover after the treatment (Fig. 1). Indeed, the irradiance (J m^{-2} s^{-1}) is crucial to preserve cells viability. In fact, using a milder irradiance reduces cells lethality and, in turn, improves the viral replication (see below), which relies on the host machinery. Indeed, the same fluence provided with two values of irradiance (0.5 or 1.0 J m^{-2} s^{-1}) led to different percentage of viability. For all the values of fluence tested, the percentage of survival cells is higher when they are treated with a lower irradiance (0.5 J m^{-}2 s^{-1}; Table 2).

A suitable UV fluence for the induction of the SSV1 replication

The effect of UV irradiation on the SSV1 replication has been evaluated by monitoring the relative amount of the SSV1 DNA in the mock-treated and UV-irradiated cultures

Table 2 Cells viability after UV treatment

Total fluence (J m^{-2})	CFU/ml Avg ± SD ($n = 3$)	Survival percentage (%)[a]
0	$8.63 \times 10^7 \pm 0.26 \times 10^7$	100
30[b]	$6.07 \times 10^7 \pm 0.30 \times 10^7$	70.33
45[c]	$3.17 \times 10^7 \pm 0.08 \times 10^7$	36.73
45[b]	$1.87 \times 10^7 \pm 0.08 \times 10^7$	21.67
60[c]	$0.60 \times 10^7 \pm 0.06 \times 10^7$	6.95
60[b]	$0.23 \times 10^7 \pm 0.02 \times 10^7$	2.67
75[c]	$0.51 \times 10^6 \pm 0.01 \times 10^6$	0.59

[a] Survival percentages calculated considering the mock-treated sample as 100 %

[b] Cells irradiated with an irradiance of 1.0 J m^{-2} s^{-1}

[c] Cells irradiated with an irradiance of 0.5 J m^{-2} s^{-1}

by semi-quantitative PCR assays. Cells were irradiated with a set of different conditions as described above and collected 8 h post-irradiation. Two single-copy genes on the host and viral chromosomes (*orc1* and *vp2*, respectively) were chosen to provide an estimation of the viral genome content under all the conditions tested. As shown, the PCR products were analysed for each sample on agarose gel at the 20th, 25th and 30th cycle of amplification (Fig. 3).

Densitometric analysis, performed by means of the software Quantity One (BioRad), revealed that the *vp2/orc1* fluorescence ratio increases in all the UV-irradiated samples compared to the mock-treated ones (when the same amplification cycle is considered). Notably, the highest value was detected for the sample treated with an irradiance of 0.5 J m^{-2} s^{-1} and a fluence of 45 J m^{-2}, which, therefore, is the most suitable condition among the several tested (Fig. 3). For this latter sample, the fluorescence intensity of *vp2* at the 20th amplification cycle is comparable to its intensity at the 25th amplification cycle of the mock-treated sample (lysogenic culture). Under the best conditions, at each amplification cycle the amount of a specific amplicon increases by a factor of 2^n, where n is the number of cycle. Since the same amount of *vp2* (fluorescence intensity) is reached 5 cycles in advance in the irradiated sample, the copy number of SSV1 in this latter is ~32-fold higher (2^5) than in the mock-treated one. Being the SSV1 copy number ~5 episomes per cell, the total viral amount reaches 160 copies per cell.

Worth of note is that, using the QIAprep Spin Miniprep Kit (Qiagen), about 20 μg of SSV1 DNA (~1.2 × 10^{12} SSV1 episomal genomes) were isolated from independent preparations, by processing 50-ml pellets of cells harvested 8 h post-irradiation (0.75 OD_{600}). If it is assumed that 1 OD_{600} culture contains about 2 × 10^8 cells/ml (as calculated by plate efficiency), 50 ml of a 0.75 OD_{600} culture contains ~7.5 × 10^9 cells. According to the data already measured by the densitometric analysis, the initial SSV1 copy number increases of 32-fold.

The SSV1 viral particles accumulate in the culture supernatant after the irradiation

Performing quantitative plaque assay to determine the SSV1 viral titre is notoriously challenging. However, we have noticed that using a not-freshly diluted culture as lawn, the resulted halos appeared clearer and easier to be counted than those obtained when a freshly diluted culture was plated. Therefore, the growth retardation induced by SSV1 infection is more pronounced when the culture is harvested from an exhausted medium.

Viral titre was determined for the culture supernatants harvested at the 8th and the 24th h post-irradiation. These two time points were chosen because after 8 h of incubation the amount of viral DNA in the cells reaches its maximum, while subsequently decreases (10–24 h), probably as consequence of the viral particles extrusion into the culture medium (our unpublished data). The highest amount of viral particles (5 10^9 PFU/ml) was produced from cells treated with a fluence of 45 J m^{-2} (irradiance of

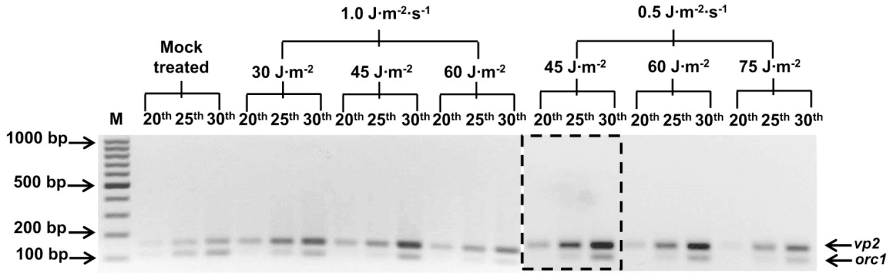


Fig. 3 Semi-quantitative PCR analysis on DNA from mock- and UV-treated SSV1 -P2 cultures. Agarose gel electrophoresis of PCR products collected at the 20th, 25th and 30th amplification cycle. *Black-straight arrows* point out to molecular weight markers as well as to host (*orc1* = 108 bp) and viral (*vp2* = 155 bp) PCR products. Densitometric analysis detected a maximum amount of SSV1 DNA for the sample irradiated with a fluence of 45 J m^{-2} and an irradiance of 0.5 J m^{-2} s^{-1} (*dashed frame*)

0.5 J m^{-2} s^{-1}) and harvested 24 h post-irradiation. Whilst, the same fluence (45 J m^{-2}), provided using an irradiance of 1.0 J m^{-2} s^{-1}, led to a lower viral titre (6 10^8 PFU/ml). This difference is due to the fact that, in the former case, a larger fraction of the cell population is viable (37 vs 22 %, Table 2), thus better supporting viral replication and virions extrusion.

Discussion

Three decades have passed since the discovery of SSV1 and it still represents a valid model to study the host–virus interaction in harsh environments (Martin et al. 1984; Ceballos et al. 2012; Schleper et al. 1992; Fröls et al. 2007a; Fusco et al. 2013). Moreover, so far it is the only member of the *Fuselloviridae* family showing a UV-inducible life cycle (Prangishvili 2013; Contursi et al. 2014a). Interestingly, the first proof of the existence of a UV-specific response in *Sulfolobus* was just derived from the transcription analysis of SSV1. In particular, the primary reaction after UV treatment of the host cells is the expression of a small transcript T_{ind}, which either acts as primer for viral replication and/or encodes for a UV-responsive transcription factor (Fröls et al. 2007a).

However, the UV irradiation exerts effect not only on the SSV1 induction but also on the host metabolism and vitality as well. Indeed, the transcriptional response in *Sulfolobus* cells is paralleled by a phase of marked growth retardation with DNA replication and cell division slowing down (Fröls et al. 2009). Exposure of cells to UV light causes the formation of two prevalent DNA lesions, i.e. cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts, the former of which has been detected in UV-irradiated cells of *S. solfataricus* (Salerno et al. 2003). The best characterized UV-damage repair system is the nucleotide excision repair (NER) pathway. Noteworthy, the genome of *S. solfataricus* encodes for homologues of the eukaryal NER system, which were found to be transcriptionally up-regulated upon UV irradiation (Salerno et al. 2003). In *Eukarya* the contribution of the NER pathway to the removal of UV-induced DNA lesions is dependent on the magnitude of the UV exposure used, i.e. on irradiance and fluence (Lee et al. 2004). Similarly, it has been shown that these two parameters represent key factors in the activation of the c-Jun N-terminal kinase (JNK), which is mediated by DNA lesions in mammalian cells. In particular, a prolonged activation of JNK was revealed when the UV fluence was administered using a lower irradiance (Adler et al. 1996). Notably, studies on UV response in *Eukarya* have evidenced that conflicting data are produced when different experimental procedures are used (Lee et al. 2004). Similar discrepancies in *S. solfataricus* (Fröls et al. 2007b, 2009; Götz et al. 2007; Salerno et al. 2003) might be attributed to the lack of a standardized protocol. Indeed, in previous literature the fluence has been only empirically determined, whist the irradiance has not been taken into account (Martin et al. 1984; Reiter et al. 1987; Schleper et al. 1992; Fröls et al. 2007a).

In this manuscript, we describe a suitable UV irradiation procedure, which is based on instrumentally measured parameters, i.e. the fluence and irradiance. A clear dose–response relationship between the UV irradiation and the host survival percentage is shown. Interestingly, cells lethality was significantly reduced through the tuning of the irradiance (0.5 or 1.0 J m^{-2} s^{-1}) (Table 2). By analogy with eukaryal systems, this effect might be due to an improved functionality and/or activation of the DNA-damage repair systems (Lee et al. 2004; Adler et al. 1996).

A fluence of 45 J m^{-2}, in combination with an irradiance of 0.5 J m^{-2} s^{-1}, turned out to be not only suitable to preserve host viability, but also to lead to the highest accumulation of viral DNA and of viral particles (Fig. 2; Table 3). Indeed, densitometric analysis of the PCR products showed an increase of the SSV1 copy number of about 32-fold, i.e. of ~160 viral genome copies per cell in the irradiated culture collected 8 h post-irradiation. Moreover, the viral titre determined under the same irradiation conditions (about

Table 3 Viral titre at 8th and 24th hours after UV treatment

Total fluence (J m^{-2})	PFU/ml 8th Avg ± SD ($n = 3$)	PFU/ml 24th Avg ± SD ($n = 3$)	PFU/ml fold of increase 8th/24th
0	$2.10 \times 10^7 \pm 0.12 \times 10^7$	$2.45 \times 10^7 \pm 0.18 \times 10^7$	1.17
30[a]	$2.03 \times 10^7 \pm 0.16 \times 10^7$	$8.26 \times 10^7 \pm 0.13 \times 10^7$	4.07
45[b]	$2.31 \times 10^8 \pm 0.25 \times 10^8$	$4.91 \times 10^9 \pm 0.11 \times 10^9$	21.26
45[a]	$3.85 \times 10^7 \pm 0.15 \times 10^7$	$6.35 \times 10^8 \pm 0.20 \times 10^8$	16.49
60[b]	$3.92 \times 10^7 \pm 0.11 \times 10^7$	$5.03 \times 10^8 \pm 0.13 \times 10^7$	12.83
60[a]	$1.76 \times 10^7 \pm 0.16 \times 10^7$	$3.68 \times 10^7 \pm 0.17 \times 10^7$	2.09
75[b]	$1.81 \times 10^7 \pm 0.08 \times 10^7$	$2.94 \times 10^7 \pm 0.16 \times 10^7$	1.62

[a] Cells irradiated with an irradiance of 1.0 J m^{-2} s^{-1}

[b] Cells irradiated with an irradiance of 0.5 J m^{-2} s^{-1}

5 10^9 PFU/ml, Table 3), was one order of magnitude higher than previously reported for *S. solfataricus* (Schleper et al. 1992). Therefore, the enhanced viability of the cell population affects viral replication and virions extrusion, thus influencing the increase of both copy number and viral titre.

Altogether these data highlight the necessity of standardizing the irradiation procedure to better compare results from different research groups.

Moreover, the establishment of a standardized protocol for SSV1 induction might have biotechnological potentialities since it allows the isolation of a huge amount of viral DNA as well as of viral particles to be employed for genetic manipulation and nanoparticles production, respectively. In this regard, viral particles from hyperthermophilic *Archaea* have been demonstrated to be exploitable for the fabrication of new nanoparticles. In particular, the rod-shaped virus SIRV2 (*Sulfolobus islandicus* rod-shaped virus 2) has been referred as a novel nanobuilding block (Evans 2009). Similarly, the virion of SSV1 can be considered, by its nature, a stable nanoparticle that is resistant to low pH and high temperature, and is, therefore, a good candidate for future biotechnological applications.

Acknowledgments Research in the authors' laboratories is supported by the grant "Programma F.A.R.O, IV tornata", founded by Università Federico II di Napoli and "Compagnia di San Paolo" (Naples Laboratory). We thank Dr Maria Ciaramella for providing us with the equipment for the UV irradiation.

References

Adler V, Polotskaya A, Kim J, Dolan L, Davis R, Pincus M, Ronai Z (1996) Dose rate and mode of exposure are key factors in JNK activation by UV irradiation. Carcinogenesis 17:2073–2076

Albers SV, Jonuscheit M et al (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Appl Environ Microbiol 72:102–111

Bolton JR, Linden KG (2003) Standardization of methods for fluence (UV Dose) determination in bench-scale UV experiments. J Environ Eng 129:209–215

Bondy-Denomy J, Davidson AR (2014) When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. J Microbiol 52:235–242

Cannio R, Contursi P, Rossi M, Bartolucci S (1998) An autonomously replicating transforming vector for *Sulfolobus solfataricus*. J Bacteriol 180:3237–3240

Cannio R, Contursi P, Rossi M, Bartolucci S (2001) Thermoadaptation of a mesophilic hygromycin B phosphotransferase by directed evolution in hyperthermophilic *Archaea*: selection of a stable genetic marker for DNA transfer into *Sulfolobus solfataricus*. Extremophiles 5:153–159

Ceballos RM, Marceau CD, Marceau JO, Morris S, Clore AJ, Stedman KM (2012) Differential virus host-ranges of the *Fuselloviridae* of hyperthermophilic *Archaea*: implications for evolution in extreme environments. Front Microbiol 3:295

Clore AJ, Stedman KM (2007) The SSV1 viral integrase is not essential. Virology 361:103–111

Contursi P, Jensen J, Aucelli T, Rossi M, Bartolucci S, She Q (2006) Characterization of the *Sulfolobus* host-SSV2 virus interaction. Extremophiles 10:615–627

Contursi P, Cannio R, Prato S, She Q, Rossi M, Bartolucci S (2007) Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. J Bacteriol 17:6339–6350

Contursi P, Cannio R, She Q (2010) Transcription termination in the plasmid/virus hybrid pSSVx from *Sulfolobus islandicus*. Extremophiles 14:453–463

Contursi P, D'Ambrosio K, Pirone L, Pedone E, Aucelli T, She Q, De Simone G, Bartolucci S (2011) C68 from the *Sulfolobus islandicus* plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. Biochem J 435:157–166

Contursi P, Fusco S, Limauro D, Fiorentino G (2013) Host and viral transcriptional regulators in *Sulfolobus*: an overview. Extremophiles 17:881–895

Contursi P, Fusco S, Cannio R, She Q (2014a) Molecular biology of fuselloviruses and their satellites. Extremophiles 3:473–489

Contursi P, Farina B, Pirone L, Fusco S, Russo L, Bartolucci S, Fattorusso R, Pedone E (2014b) Structural and functional studies of Stf76 from the *Sulfolobus islandicus* plasmid-virus pSSVx: a novel peculiar member of the winged helix-turn-helix transcription factor family. Acids Res, Nucl. doi:10.1093/nar/gku215

Eilers BJ, Young MJ, Lawrence CM (2012) The structure of an archaeal viral integrase reveals an evolutionarily conserved catalytic core yet supports a mechanism of DNA cleavage in trans. J Virol 86:8309–8313

Evans DJ (2009) Exploitation of plant and archaeal viruses in bionanotechnology. Biochem Soc Trans 37:665–670

Fröls S, Gordon PM, Panlilioc MA, Schleper C, Sensen CW (2007a) Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology 365:48–59

Fröls S, Gordon PM, Panlilio MA, Duggin IG, Bell SD, Sensen CW, Schleper C (2007b) Response of the hyperthermophilic archaeon *Sulfolobus solfataricus* to UV damage. J Bacteriol 189:8708–8718

Fröls S, White MF, Schleper C (2009) Reactions to UV damage in the model archaeon *Sulfolobus solfataricus*. Biochem Soc Trans 37:36–41

Fusco S, She Q, Bartolucci S, Contursi P (2013) T(lys), a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. J Virol 87:5926–5936

Götz D, Paytubi S, Munro S, Lundgren M, Bernander R, White MF (2007) Responses of hyperthermophilic crenarchaea to UV irradiation. Genome Biol 8:R220

Iverson E, Stedman K (2012) A genetic study of SSV1, the prototypical fusellovirus. Front Microbiol 3:200

Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C (2003) A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. Mol Microbiol 48:1241–1252

Kraft P, Kummel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM (2004a) Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol 78:7438–7442

Kraft P, Oeckinghaus A, Kummel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM (2004b) Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA-binding protein. J Virol 78:11544–11550

Lee DF, Drouin R, Pitsikas P, Rainbow AJ (2004) Detection of an involvement of the human mismatch repair genes *hMLH1* and *hMSH2* in nucleotide excision repair is dependent on UVC fluence to cells. Cancer Res 64:3865–3870

Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W (1984) SAV1, a temperate UV-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J 3:2165–2168

Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzerd E, Young MJ, Bothner B, Lawrence CM (2008) Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology 376:270–278

Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology 185:242–250

Prangishvili D (2013) The wonderful world of archaeal viruses. Annu Rev Microbiol 8:565–585

Redder P, Peng X, Brugger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ Microbiol 11:2849–2862

Reiter WD, Palm P, Yeats S, Zillig W (1987) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol Gen Genet 209:270–275

Reiter WD, Palm P, Zillig W (1988) Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res 16:1–19

Salerno V, Napoli A, White MF, Rossi M, Ciaramella M (2003) Transcriptional response to DNA damage in the archaeon *Sulfolobus solfataricus*. Nucleic Acids Res 31:6127–6138

Schleper C, Kubo K, Zillig W (1992) The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sci USA 89:7645–7649

Snyder JC, Stedman K, Rice G, Wiedenheft B, Spuhler J, Young MJ (2003) Viruses of hyperthermophilic *Archaea*. Res Microbiol 154:474–482

Stedman KM, Schleper C, Rumpf E, Zillig W (1999) Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. Genetics 152:1397–1405

Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res Microbiol 154:295–302

Wiedenheft B, Stedman KM, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M (2004) Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J Virol 78:1954–1961

Wolin EA, Wolin MG, Wolfe RS (1963) Formation of methane by bacterial extracts. J Biol Chem 238:2882–2886

Yeats S, McWilliam P, Zillig W (1982) A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. EMBO J 1:1035–1038

Zhan Z, Ouyang S, Liang W, Zhang Z, Liu ZJ, Huang L (2012) Structural and functional characterization of the C-terminal catalytic domain of SSV1 integrase. Acta Cryst D68:659–670

Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I (1980) The *Sulfolobus*-"Caldariella" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch Microbiol 125:259–269

2.4 Unravelling the role of the F55 regulator in the transition from lysogeny to UV induction of *Sulfolobus* spindle-shaped virus 1

Salvatore Fusco[a], Qunxin She[b], Gabriella Fiorentino[a], Simonetta Bartolucci[a] and Patrizia Contursi[a#]

Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, 80126 Napoli, Italy[a]; Danish Archaea Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark[b]

Running Head: Transition from lysogeny to UV-induction of SSV1

[#] Address correspondence to Patrizia Contursi, contursi@unina.it

Keywords: Lysogeny regulator; Molecular switch; *Sulfolobus* spindle-shaped virus; UV induction; SSV1; *Fuselloviridae*

Submitted to Journal of Virology

Abstract

Sulfolobus spindle-shaped virus 1 represents a model for studying virus-host interaction in harsh environments and it is so far the only member of the *Fuselloviridae* family that shows a UV-inducible life cycle. Although the virus has been extensively studied, mechanisms underpinning the maintenance of lysogeny as well as those regulating the UV induction have received little attention. Recently, a novel SSV1 transcription factor, F55, was identified. This factor was able to bind *in vitro* to several sequences derived from the early and UV-inducible promoters of the SSV1 genome. The location of these binding sites together with the differential affinity of F55 towards these sequences, led to the hypothesis that this protein might be involved in the maintenance of the SSV1 lysogeny. Herein, it is reported an *in vivo* survey of the molecular events occurring at the UV-inducible region of the SSV1 genome, with a focus on the binding profile of F55 before and after the UV irradiation. The binding of F55 to the target promoters correlates with transcription repression, whereas its dissociation is paralleled by transcription activation. Therefore, we propose that F55 acts as molecular switch for the transcriptional regulation of the early viral genes.

Importance

Functional genomic studies of SSV1 proteins have been hindered by the lack of similarity with other characterized proteins. As a result, only few insights into their *in vivo* roles have been gained throughout the last three decades. Herein, we report the first *in vivo* investigation of a SSV1 transcription regulator, F55, that exerts a key role in the transition from the lysogenic to the induced state of SSV1. We show that F55 regulates the expression of the UV-inducible as well as of the early genes. Moreover, the differential affinity of this transcription factor towards these targets allows a fine-tuned and temporal coordinated regulation of transcription of viral genes.

Introduction

The majority of viruses isolated from *Archaea*, the third domain of life, show virion morphotypes that have not been observed for viruses infecting *Bacteria* and *Eukarya*. Consequently, eight novel viral families have been introduced in order to classify these novel viruses, i.e. *Fuselloviridae*, *Lipothrixviridae*, *Rudiviridae*, *Guttaviridae*, *Globuloviridae*, *Bicaudaviridae*, *Ampullaviridae* and *Clavaviridae* but so far, there are still several unique archaeal viruses that remain to be classified (1-3). Many analyses of environmental samples have shown that spindle-shaped viruses are abundant and occupy several niches, including deep sea hydrothermal vents (4, 5), hypersaline environments (6,7), anoxic freshwaters (8), cold Antarctic lakes (9), terrestrial hot springs (10-12) and acidic mines (13, 14), where they frequently outnumber head-tailed viruses. Notably, this unique virion morphotype seems to be a hallmark of viruses infecting *Archaea*, since it has never been observed for bacteriophages and eukaryal viruses (15). To date, the family of *Fuselloviridae* comprises nine members (SSV1, SSV2, SSV4, SSV5, SSV6, SSV7, SSV8, SSV9 and ASV1) isolated from several geographic locations (2). The lack of structural and functional characterization of proteins encoded by the fuselloviral genomes has limited the dissection of fundamental processes, such as: i) virion uptake, assembly and release, ii) transcriptional regulation, iii) genome replication and iv)

lysogeny/induction switch. Indeed, with the exception of few proteins (16-23), structural and functional annotations in main databases are not available for fuselloviral gene products (2, 22).

SSV1 is the most extensively characterized member of this viral family and is the only one showing a UV-inducible life cycle. Upon infecting a host cell, SSV1 integrates one copy of its genome into the host chromosome at an arginyl-tRNA gene, forming a provirus (24). However, unlike the well characterized lambda phage, for which lysogenic cells harbour only the provirus, SSV1 lysogens carry also ~5 copies of the episomal DNA per cell. This led to the hypothesis that SSV1 expresses a minimal set of genes to ensure a basal level of replication that is required for maintaining the carrier stage (23). Moreover, structural proteins (VP1, VP2 and VP3) are constitutively expressed under conditions for which viral replication is not induced (23), allowing the production of viral particles by the lysogenic cells. Upon exposure to UV light, SSV1 exhibits a temporally coordinated pattern of gene expression. At first, it activates the expression of a UV-inducible transcript (T_{ind}), followed by the transcription of the early (T_5, T_6 and T_9), late ($T_{1/2}$, T_3, T_x, and $T_{4/7}$) and late-extended ($T_{4/7/8}$) RNAs. This cascade of events leads, in turn, to the induction of the SSV1 genome replication and eventually to a steep increase of the viral titre (25).

Despite the extensive characterization that SSV1 has received over the last 30 years (25-28), mechanisms underpinning the transition from the lysogenic growth to the viral induction are still murky. Very recently, a new mRNA (T_{lys}) was discovered, which was transcribed in the opposite direction of the UV-inducible T_{ind} (23). Since T_{lys} is one of the most abundant SSV1 transcripts during lysogenic growth, it was suggested that the encoded protein could play a fundamental role in the maintenance of the carrier stage. Indeed, T_{lys} encodes for a 55-amino acids protein (F55) that interacts, in a concentration-dependent manner, with tandem-repeated sequences clustered within the UV-inducible region of the viral genome. Moreover, F55 could act as a transcription repressor since these operators encompass both the transcription start sites (TSSs) and the B recognition elements (BREs) of the T_5, T_6, T_{ind} and T_{lys} promoters (23). So far, F55 is the only transcription repressor for which a defined role in the regulation of a fuselloviral life cycle has been proposed. Herein, we report an *in vivo* survey of the molecular events occurring upon irradiation at the UV-inducible region of the SSV1 genome, with a focus on the pleiotropic effect of F55 on several SSV1 promoters. We show that F55 acts as the molecular switch controlling the SSV1 life cycle.

Materials and methods

Strains, media, growth conditions and UV irradiation

A SSV1 lysogenic strain of *Sulfolobus solfataricus* (SSV1-InF1) was generated, by infecting the uracil auxotrophic mutant InF1 (29) as described elsewhere (30). An aliquot from a frozen sample of SSV1-InF1 culture was thawed and a few microliters were spotted onto an SCVYU-Gelrite plate and incubated at 75°C. After 3-5 days of incubation, local growth-areas (spots) were inoculated into 50 ml of SCVYU medium, i.e. a glycine-buffered Brock's basal salt solution, supplemented with 0.2% sucrose weight volume^{-1} (wt vol^{-1}), 0.2% casamino acids (wt vol^{-1}), 1× vitamins (31), 0.005% yeast extract (wt vol^{-1}) and uracil 0.02 mg ml^{-1}, with pH adjusted to 3.5 with concentrated H_2SO_4. Cultivation of the *Sulfolobus* strains was conducted in a 250-ml Erlenmeyer flask with a long neck at 75°C with a shaking rate of 180 rpm using a MaxQ™ 4000 Benchtop Orbital Shakers (Thermo Scientific). Cell growth was monitored spectrophotometrically at 600 nm (OD_{600}) by means of a Variant Cary® 50

Bio UV/Visible Spectrophotometer (McKinley Scientific). Once the culture reached the logarithmic phase of growth, it was diluted to an optical density value of 0.05 OD_{600} in 50 ml of fresh SCVYU medium and let to grow up to 0.5-0.8 OD_{600}.

The SSV1-InF1 culture was UV-irradiated by using a fluence of 45 J m^{-2} and an irradiance of 0.5 J m^{-2} s^{-1}, as described elsewhere (32). Growth was spectrophotometrically monitored before and after the treatment to construct comparative growth curves. Samples were taken after 2, 4, 6, 8 and 10 hours post-treatment and pellets for each time point were obtained through centrifugation at 3,000 × *g* for 10 min using the Centrifuge 5810R (Eppendorf). Pellets were treated for total DNA, RNA and protein preparations (see below).

Semi-quantitative PCR and *EcoRI* total DNA digestions

SSV1-InF1 cell pellets, obtained as described previously, were treated for highly pure total DNA preparation using the DNeasy tissue kit (Qiagen), following the manufacturer's instructions. The concentration of the DNA samples was spectrophotometrically determined using a Nanodrop 2000 Spectrophotometer (Thermo-Scientific), by performing the measurements in triplicates. Their purity was assessed by the 260nm/280nm adsorption ratio and only DNA samples with a value of 1.8 or higher were used for subsequent experiments.

In order to detect variations of the viral DNA content among the different samples, total DNA samples from irradiated SSV1-InF1 cultures were analysed by semi-quantitative PCR (sqPCR). With this aim, two primer couples were designed using Primer3 software (available at the website: http://bioinfo.ut.ee/primer3-0.4.0/), in order to amplify: (i) a 155-bp region of the SSV1 single-copy gene *vp2* and (ii) a 108-bp region of the host single-copy gene *orc1* (Table 1).

Table 1
Primers used for the semi-quantitative PCR assays on ChIP samples

Name	Sequence (5'-3')	Length (nt)
T_5-fw	CCCAAACACTGTGTATATAGAG	22
T_5-rv	AGTTTGTGCCATATTCCCAT	20
T_6-fw	ATGATAATATTAAATGATTCACGAT	25
T_6-rv	TTTCGGGTTTGGGGTGAAAC	20
T_{ind}-fw	CTGCTGTCTGACAAGAGTTT	20
T_{ind}-rv	GATTTTGCACATCCCATATT	20
T_{lys}-fw	ATCGTGAATCATTTAATATTATCAT	25
T_{lys}-rv	ATTGGAATCGAAACGGTCAC	20
orc1-fw	TATAAATTGTTATAGACATAGAACGCTGTA	30
orc1-rv	TTAAATACTTCTTGTGCCGATAGTCC	26
vp2-fw	GGAGGGTACATCGCTACCTTATGA	24
vp2-rv	CAGTAGGGCTGACAGTAAACTACG	24

A sqPCR master mix was set up as follows: 1× Taq buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP mix, 0.6 µM *orc1*-fw, 0.6 µM *orc1*-rv, 0.6 µM *vp2*-fw, 0.6 µM *vp2*-rv and 50U ml^{-1} of Taq DNA Polymerase (Thermo Scientific) in a total volume of 500 µl. The master mix was then separated into seven aliquots of 60 µl each in different tubes and labelled as control sample (mock-treated collected at 0.5 OD_{600}), 2h, 4h, 6h, 8h and 10h (UV-treated samples). To each tube, 100 ng of total DNA from mock-treated or UV-irradiated samples, were added as template. The aliquots were then split up into three sub-aliquots (20 µl each), labelled (20^{th}-25^{th}-30^{th}) and placed into a Mastercycler Personal (Eppendorf®). A negative (no template) and two positive controls of the amplification conditions were also included in the analysis. The thermal cycling protocol was as follows: an initial denaturation step of 5 min at 95°C, followed by 30 cycles of 40 sec at 95°C, 40 sec at 62°C, and 1 min at 72°C. A 5 final step at 72°C has been carried out for 10 min at the end of the 30^{th} cycle. For each time point, tubes were taken from the thermocycler at the 20^{th}, 25^{th} and 30^{th} cycle of amplification. PCR products were run a 2% agarose gel (wt vol^{-1}) in 1× TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA). The same DNA samples were enzymatically digested using the sequence-specific endonucleases *EcoRI*, whose target sequence recurs four times in the SSV1 genome. In brief, 2 µg of each total DNA sample were digested overnight with 20 units of *EcoRI* (Roche) at 37°C and run on a 1% agarose gel (wt vol^{-1}).

Northern blot analysis

High pure total RNA samples were prepared from mock-treated and UV-irradiated SSV1-InF1 cultures by means of the TRIzol reagent (Sigma Aldrich®), according to the manufacturer's instructions. RNA pellets were dissolved in nuclease-free water and concentration was determined by means of a Nanodrop 1000 Spectrophotometer (Thermo Scientific). Samples purity was checked through the determination of the 260nm/280nm adsorption ratio and only samples showing a ratio of 1.9 or higher were used for subsequent experiments.

For each sample, 20 µg of total RNA were run on a denaturing, formaldehyde-containing 2.0% agarose gel (wt vol^{-1}) and then transferred onto a nylon membrane (Hybond-XL; Amersham-Pharmacia). T4 polynucleotide kinase (Fermentas Life Sciences) was used to label 5' ends of single-stranded oligonucleotides T_{lys}-rv (5'-AAGTTCTTCAATGCGTCTTCTGATT-3'), T_{ind}-fw (5'- TCTGAGCTACTAATACTGCTT GAAT-3') and 16S-rv (5'-CTCTCCTACTCGGGTGGAGCAAC) with radioactive [γ-32] ATP, following the manufacturer's instructions. Purification of radiolabeled oligonucleotides was achieved by gel filtration chromatography using Illustra Nick columns (Amersham Biosciences) and hybridizations with single-stranded DNA probes were carried out as described elsewhere (33). Probes were eventually removed by boiling in 0.1% (wt vol^{-1}) sodium dodecyl sulphate (SDS) for 10 min in order to reuse the membrane for subsequent hybridizations. Radioactive signals were quantified by means of a Molecular Dynamics Bio-Rad PhosphorImager (Quantity One software) to determine the relative abundance of the transcripts (16S, T_{lys} and T_{ind}). Determination of T_{lys}/16S and the T_{ind}/16S ratio was performed to normalize RNA signals and compensate for operator errors. The size of the Northern blot signals was determined using the RNA molecular weight markers (Roche) as standards.

Western blot analysis

Cells from 30 ml of mock-treated and UV-irradiated SSV1-InF1 cultures were harvested by centrifugation, and pellets were suspended in 2 ml of 10 mM Tris-HCl pH 8.0 (lysis buffer), containing Complete 600 protease inhibitor cocktail tablets (Roche). Cells were lysed by sonication at 20% of the maximal amplitude for 3 min, alternating 20s of pulse-on and 20s of pulse-off, by means of an ultrasonic liquid processor (Heat System Ultrasonic, Inc.). Lysates were centrifuged at 30,000 × *g* (SW41 rotor; Beckman) for 30 min in order to clarify crude protein extracts. Protein concentration was spectrophotometrically determined by standardized Coomassie (Bradford) Protein Assays and normalized for each sample, as described elsewhere (34). In brief, Bradford assays were performed on bovine serum albumin as standard (BSA, Thermo Scientific) to construct a calibration curve. In particular, known protein concentrations (1, 2, 4, 6, 8 μg/ml) were plotted against their absorbance at 595nm. Concentration of protein samples was determined by means of the standard curve and samples quality assessed by SDS-PAGE followed by Coomassie staining.

A serum sample (15 mg of total protein) from an F55-immunized rabbit (Innovagen AB) was loaded on a 1-ml HiTrap Protein A column (GE Healthcare) connected to a fast-performance liquid chromatography system (ÄKTA, GE Healthcare) and total IgGs were purified following the manufacturer's instructions. Antibodies-containing fractions were pooled, dialyzed against 20 mM sodium phosphate buffer pH 8.0 and protein concentration (0.4 mg ml^{-1}) determined by the Coomassie (Bradford) Protein Assay, as described above. Antibodies integrity was checked by running 10 μg of total IgG sample on 15% SDS-PAGE. For western blot hybridizations, total protein samples (10 μg) were run on 15% SDS-PAGE and electro-transferred onto an Immobilon-PVDF membranes (Millipore). Subsequently, membranes were treated as follows: i) incubated for 1 hours at room temperature in blocking solution, i.e. 1× TBS-T (50 mM Tris pH 7.5, 150 mM NaCl and 7 0.1% Tween-20) containing 5% BSA (wt vol^{-1}) (Sigma); ii) incubated for 16 hours at 4°C with total IgG sample (see above) diluted (1:1000) in blocking solution; iii) washed three times for 15 min with 1× TBS-T at room temperature; iv) incubated for 1 hour at room temperature with horseradish peroxidase conjugated goat anti-rabbit IgG (Roche) diluted (1:10000) in 1× TBS-T and finally v) washed twice for 15 min with TBS-T and vi) washed once with TBS. Detection by enzyme-linked chemiluminescence was performed by means of the kit ImmobilonTM Western Chemiluminescent HRP Substrate (Millipore) and using a ChemiDoc™ XRS+ System (BioRad), according to the manufacturer's instructions. The F55 concentration in the analysed samples was determined using a calibration curve, which was constructed by plotting known amounts of F55 (from 10 to 35 ng) against densitometric values measured using the Quantity One software (BioRad). The Western blot data were validated by repeating the experiments in triplicates.

Chromatin immunoprecipitation assay and semi-quantitative PCR (ChIP-sqPCR analysis)

SSV1-InF1 cells were cultivated in 900 ml of SCVYU medium into a 2-L Erlenmeyer flask and grown to 0.6 OD_{600}. A 300-ml control sample was harvested before UV-treating the remaining volume (600 ml). Samples were harvested after 2 and 4 hour of incubation and rapidly cooled down at 37°C, before adding 1% formaldehyde (wt vol^{-1}) (Sigma) for DNA-proteins crosslinking. Incubation at 37°C was performed for 5 min with a shaking rate of 120 rpm. The crosslinking reaction was then quenched by adding glycine (125 mM final concentration) to the cultures, followed by an additional

incubation of 5 min at 37°C. Afterwards, cells were collected by centrifugation at 3,000 × *g* for 15 min (JA-14 rotor; Beckman). Pellets were washed twice with 1× PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4 • 2 H_2O, 1.76 mM KH_2PO_4, pH 7.4) and finally resuspended in 6 ml of 1× DNase Buffer (400 mM Tris-HCl, 100mM NaCl, 60 mM MgCl2 and 10 mM 8 $CaCl_2$, pH 7.9). Cells were lysed by sonication at 20% of the maximal amplitude for 18 min, alternating 3 s of pulse-on and 9 s of pulse-off, by means of an ultrasonic liquid processor (Heat System Ultrasonic, Inc.). Lysates containing cross-linked DNA-proteins complexes were clarified by centrifuging for 30 min at 30,000 × *g* (SW41 rotor; Beckman). In order to further narrow down the length of the DNA fragments, DNase I RNase free (Roche) was added at the final concentration of 30U/ml, followed by incubation of 30 min at 37°C. To recover DNA-protein complexes, samples were subjected to phenol extraction, by adding 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) and centrifuging for 10 min at 3,000 × *g*. The DNA-protein complexes were recovered by ethanol precipitation and centrifugation for 10 min at 15,000 × *g*. Pellets were dissolved in 20 mM sodium phosphate buffer pH 8.0 and treated with 100 µg ml^{-1} of RNase A (Invitrogen) for 1 h at 37°C. Cross-linked F55-DNA complexes were enriched by affinity chromatography using the 1-ml HiTrap Protein A column (GE Healthcare) connected to a fast-performance liquid chromatography system (ÄKTA). In brief, the column was equilibrated in binding buffer (20 mM sodium phosphate buffer pH 8.0), loaded with purified IgG from F55-immunized rabbit (Innovagen AB) and washed with 5 ml of binding buffer. DNA-protein samples were loaded onto the column and 5 ml of binding buffer were used to elute unspecific complexes. F55-DNA specific complexes were then eluted from the column by washing with 0.1 M citrate buffer pH 6.2 (elution buffer). In order to de-crosslink F55-DNA complexes, the eluted fractions were pooled and incubated at 65°C for 16 hours with a shaking rate of 180 rpm. Free-protein DNA was obtained by purification using the MinElute PCR purification kit (Qiagen), according to the manufacturer's instructions. The concentration of the DNA samples was measured using a Nanodrop 2000 Spectrophotometer, as described above. To check the enrichment of DNA fragments containing F55 target sequences, semi-quantitative PCR reactions (ChIP-sqPCR) were set up using the primer couples listed in Table 1. Master mixes were prepared as follows: 1× Taq Buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP mix, 0.3 µM primer-fw, 0.3 µM primer-rv, and 50U ml^{-1} of Taq DNA Polymerase (Thermo Scientific) in a total volume of 250 µl. Two aliquots of 50 µl each were used to perform negative and positive controls of the amplification, whilst to the remaining 150 µl was added 1 ng of enriched DNA obtained by ChIP (see above). The aliquot of 150 µl was then spit up into three tubes (50 µl each), labelled (20^{th}-25^{th}-30^{th}) and placed into a Mastercycler Personal (Eppendorf®). The thermal cycling protocol was as follows: an initial denaturation step of 3 min at 95°C, followed by 30 cycles of 40 sec at 95°C, 40 sec at 50°C, and 1 min at 72°C. For amplification of the host sequence *orc1* the annealing temperature was 62°C. Tubes were taken from the thermocycler at the 20^{th}, 25^{th} and 30^{th} cycle of amplification. PCR products were run on a 2% agarose gel (wt vol^{-1}) in 1× TAE buffer. The same procedure has been carried out by using as templates the enriched DNA samples from mock-treated and UV-irradiated SSV1-InF1 cells collected after 2 and 4 hours post-irradiation.

Results

Host response to the UV irradiation

In order to investigate the role of F55 in the SSV1 life cycle, a *S. solfataricus* InF1 strain carrying the virus (SSV1-InF1) was treated by UV irradiation as described elsewhere (32). The effect of the treatment on its physiology was assessed by measuring the generation time and cell viability of the mock- and UV-treated cultures. We found that, in the immediate aftermath of the treatment, cell viability of the UV-treated culture was approximately 35% of that of the untreated control. Furthermore, the growth rate of the UV-irradiated sample was slower than that of the mock-treated control, probably as a consequence of the stress caused by both the irradiation itself and the viral induction (Fig. 1), as also observed in previous experiments (32).

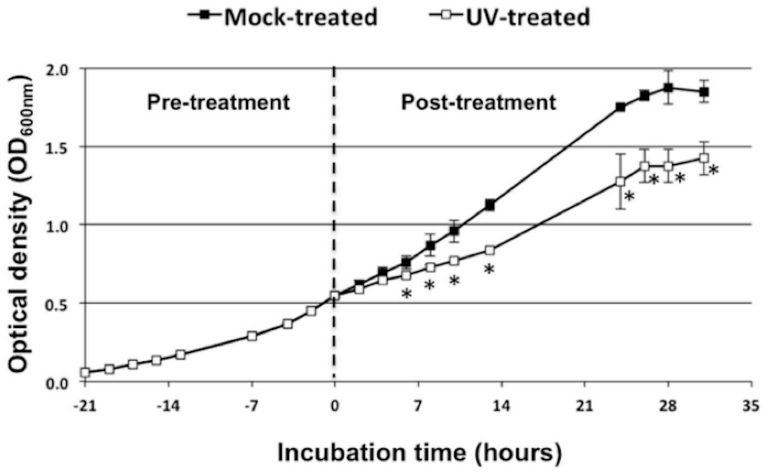


Figure 1. Growth curves of mock- and UV-treated SSV1-InF1 cultures. The OD_{600nm} values were measured every two hours and plotted versus the incubation time. Before being mock- or UV-treated, cells were grown exponentially until 0.5 OD_{600nm} (21^{th} hour of incubation). Afterwards, the culture was split in two halves and incubated back to 75˚C to estimate the effect of the treatment. Error bars show the standard deviation of experimental point (n = 3) * *p value* < 0.05.

Time-course of the SSV1 replication induction upon UV irradiation

Induction of SSV1 replication upon UV irradiation was investigated in a time course experiment of a 10-hours window, following an UV irradiation protocol reported recently (32). Samples were withdrawn from both mock-treated and UV-irradiated SSV1-InF1 cultures every two hours during incubation and the relative amount of SSV1 was estimated using semi-quantitative PCR analysis of total DNAs extracted from cell samples. Two single copy genes were chosen for this analysis, i.e. the viral *vp2* representing the SSV1 genome and *orc1* for the host genome. PCR was conducted for 20, 25 and 30 cycles of amplification and the yielded PCR products were analysed by agarose gel electrophoresis. We found that, when amplified for the same numbers of PCR cycle, the intensity of the PCR product of the internal control *orc1* (108 bp) was very similar in all the samples. However, the intensity of the *vp2* amplicon (155 bp) is significantly increased in the samples after UV-irradiation, starting from hour 2 and peaking at hour 8 post-irradiation (Fig. 2). We reasoned that the increase of viral DNA, before completion of viral gene expression (Fig. 2, 2h post-treatment), resulted from the irradiation-mediated cell division arrest (Fig. 1) that did not affect the basal viral replication. Thereafter, transcriptional activation was

completed for all viral genes (25) and the induction of the SSV1 replication accounted for the further increase of relative amounts of viral DNA in UV-irradiated cultures (Fig. 2, 6-10 hours)

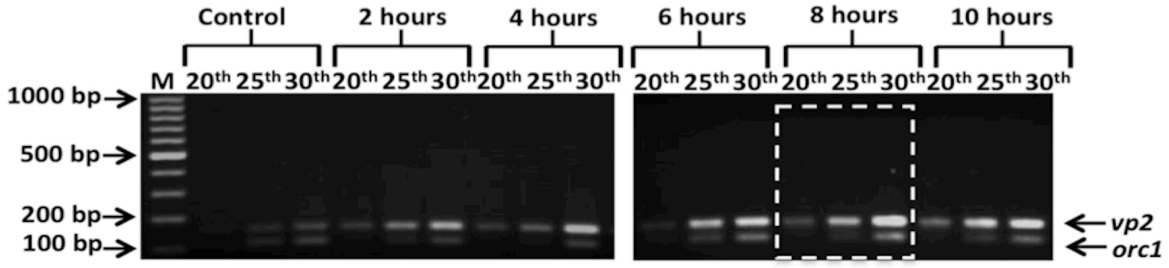


Figure 2. Time course of the viral replication induction after UV irradiation by semi-quantitative PCR. Black-straight arrows point out to molecular weight markers as well as to the host (*orc1* = 108 bp) and the viral (*vp2* = 155 bp) PCR products. Total DNA samples were prepared from mock-treated cells (control) and UV-treated cultures collected at 2, 4, 6, 8 and 10 hours post irradiation. The maximum amount of SSV1 DNA is detectable for the sample collected 8 hours post treatment (dashed white box).

To confirm the temporal onset of UV-induced viral replication, the total DNA extracted was digested with *EcoRI* and analysed by agarose gel electrophoresis. As shown in figure 3, the *EcoRI* fragments of the SSV1 genome (7.9, 2.9, 2.4 and 2.1 Kb), become dominant in the UV-irradiated cells with the highest amount detectable 8-10 hours post-irradiation. Altogether, the above data indicated that the previously established procedure for UV irradiation was suitable for studying the induction of SSV1 replication (32).

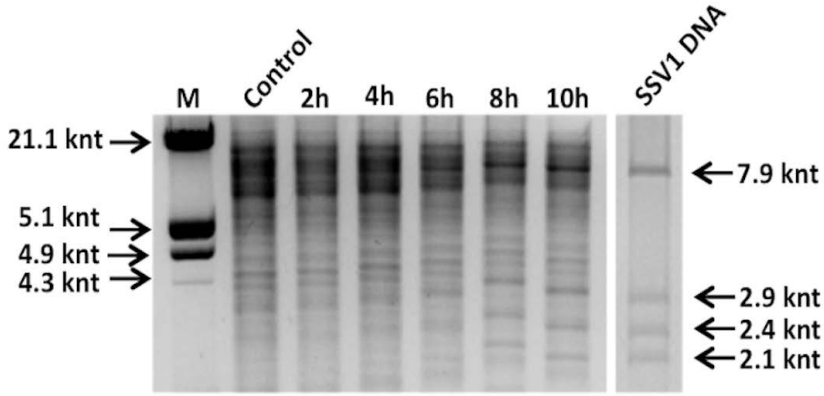


Figure 3. Detection of the viral DNA after UV irradiation by *EcoRI* restriction analysis.
The restriction profile of total DNA samples from control culture, SSV1-InF1 irradiated cells as well as from SSV1 episomal DNA digested with *EcoRI*, is shown. Black straight arrows point out to the molecular weight markers and to the SSV1-derivated fragments (7.9, 2.9, 2.4 and 2.1 Kbp). The intensity of the SSV1 fragments is higher for the sample collected 8-10 hours post irradiation.

T_{ind} and T_{lys} transcription patterns in mock- and UV-treated cells

In order to reveal the molecular circuit regulating the maintenance of SSV1 lysogeny and to identify the molecular switch to the induced state of SSV1, a transcriptional analysis was carried out on the head-to-head oriented transcripts T_{ind} and T_{lys}, the main players of the SSV1 UV-inducible expression (Fig. 4A). Total RNAs samples were prepared from mock- and UV-treated SSV1-InF1 cells and analysed by Northern hybridization to detect T_{ind} and T_{lys} transcripts (Fig. 4B). Normalization for

all samples was performed by measuring the signal of the housekeeping gene 16S and by calculating the T_{ind}/16S and T_{lys}/16S signals ratio.

T_{ind} was not detectable in mock-treated cells, which is consistent with the transcript being repressed in the carrier state (lysogeny). Conversely, soon after UV irradiation (2 hours post-treatment) two typical T_{ind} hybridization signals (25) were detected (200 nt and 300 nt). Then, the T_{ind}/16S ratio increased by 3-fold 4-6 hours after UV irradiation. Thereafter, the intensity of the T_{ind} signal dropped sharply, and became undetectable 6 hours post-treatment (Fig. 4B).

On the other hand, the antagonistic transcript T_{lys} showed a down-regulation of its expression in the mock-treated culture between 4 and 8 hours, as judged by the 50% decrease of the signal ratio T_{lys}/16S. Interestingly, this value did not change between the 2nd and the 4th hour post-irradiation (Fig. 4B), despite the increase of the SSV1 copy number in the same time window, as evidenced by sqPCR (Fig. 2).This suggests that there could exist a fine-tuning mechanism regulating the expression of these two convergent genes. Furthermore, the 300-nt T_{lys} signal did not occur as a discrete band but formed a continuous smearing with smaller RNA products mainly visible between 8 -10 hours post-treatment (Fig. 4B). This indicated that T_{lys} could be degraded by a post-transcriptional regulation mechanism. If so, the apparent increase of T_{lys} signal should not reflect the actual change of content of the encoded transcriptional regulator (F55) in the UV-irradiated cells (see below).

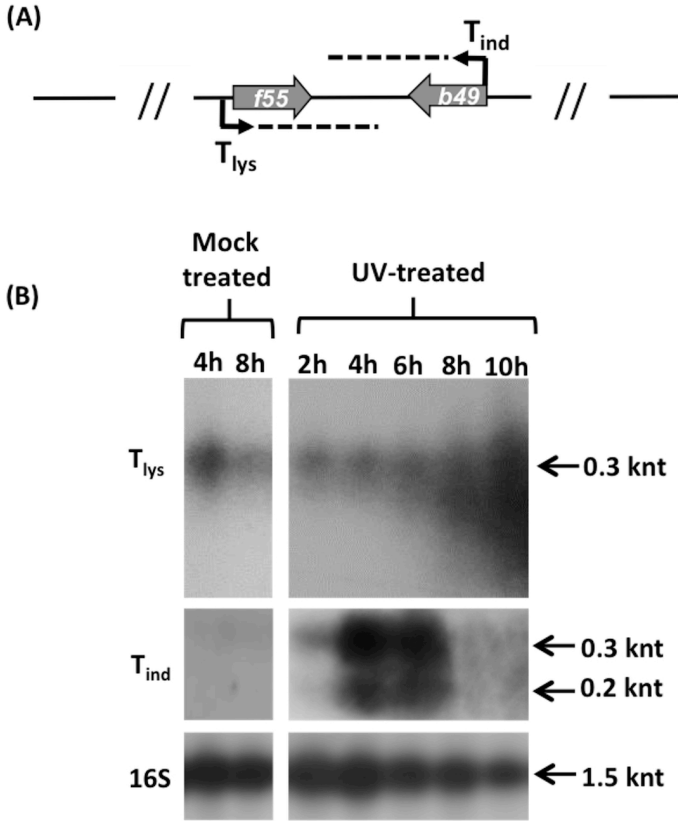


Figure 4. Transcription analysis of T_{lys} and T_{ind}. (A) Schematic representation of the head-to-head oriented transcripts T_{ind} and T_{lys}. **(B)** SSV1-InF1 total RNAs isolated from mock- and UV-treated cells at different time points were analyzed by Northern hybridization to detect T_{lys} and T_{ind} transcripts. The hybridization signals were normalized using the housekeeping gene *16S*. The T_{ind} transcript was expressed over a short time window and only in UV-irradiated cells whereas T_{lys} was detectable in both mock-treated and UV-irradiated cells. ~~A~~ Degradation occurs around the 8-10th hour only for the T_{lys} transcript.

In vivo expression and DNA-binding activity of F55

To determine the representatively of the putative lysogeny regulator F55 after UV irradiation, the content of F55 in UV-induced and mock-treated cells was analysed by densitometric analysis on immunoblots (Fig. 5A) and compared to known amounts of recombinant F55 as a reference (Fig. 5B). The intensity of the chemiluminescent signal obtained from the standards was linear in the range considered. Interestingly, throughout the whole time window analysed, the concentration of F55 in UV-treated cells was lower than that in the corresponding mock-treated samples (Fig. 5C). Furthermore, a 50% drop of the F55 cellular concentration (Fig. 5C, 2h grey bar) coincided with the onset of the T_{ind} transcription (Fig. 4B, lanes 2h UV-treated). Together these results support the hypothesis that F55 plays a role in maintaining the lysogeny. Furthermore, the simultaneous reduction of F55 protein and the augmentation of T_{lys} signal at 8^{th}-10^{th} hour (Fig. 4B) suggest that the regulation of T_{lys} expression must occur at the post-transcriptional level.

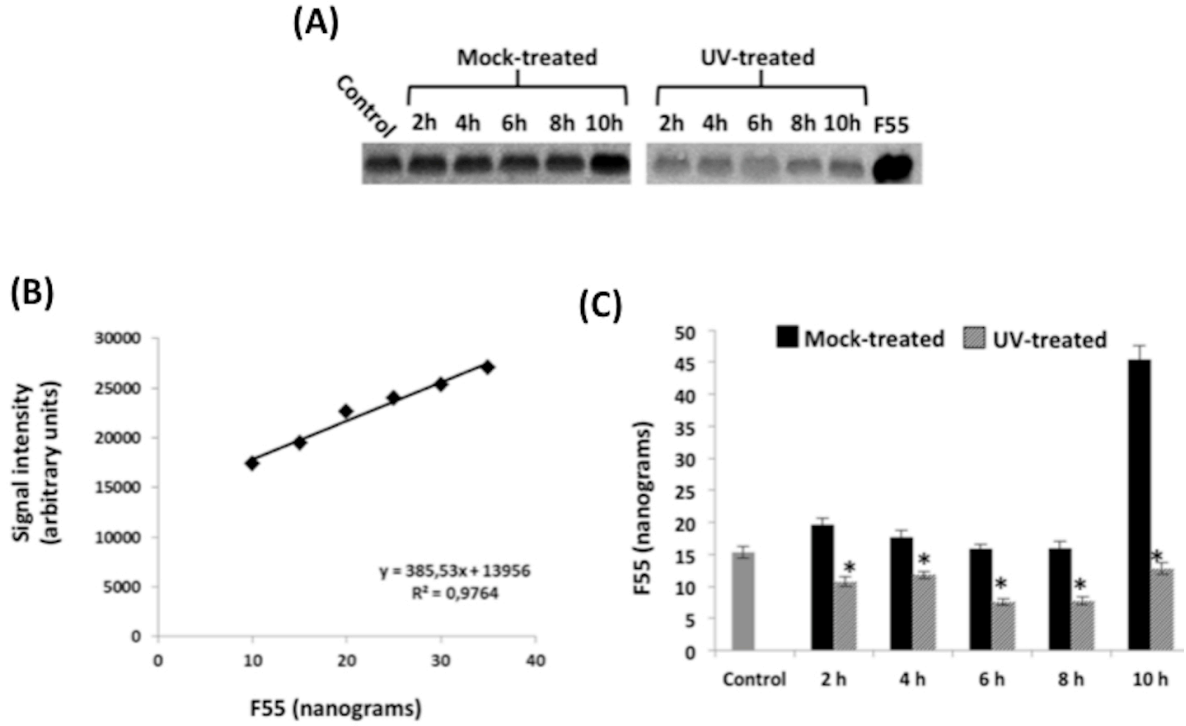


Figure 5. Western blot analysis of F55 in mock- and UV-treated InF1-SSV1 cells. (A) Western blot of cell extracts from control, mock- and UV-treated samples. **(B)** F55 quantification has been performed using a standard curve. **(C)** Quantitative data are reported as histograms. Error bars show the standard deviation of experimental point (n = 3) and a control sample (grey bar) is used as reference. The amount of F55 in the UV-treated culture is significantly lower than the mock one at every time point analysed (*p *value* < 0.05).

It was previously demonstrated that F55 binds *in vitro*, with differential affinity, to DNA sequences derived from the promoters of T_5, T_6, T_{ind} and T_{lys} transcripts. Based on the location of the binding sites, i.e. partially overlapping the BRE or encompassing the TSS, it was hypothesized that this regulator could function as transcriptional repressor (23). To investigate binding of F55 to the aforementioned

promoters *in vivo* (Fig. 5A), DNAs were recovered by ChIP assays before and after the UV treatment and subjected to sqPCR analysis (ChIP-sqPCR) as described above. In lysogenic cells, promoter regions of T_5, T_6, T_{ind} and T_{lys} were detected in ChIP-sqPCR already after 25 cycles (Fig. 6B, control panel), indicating that F55 binds to these promoters *in vivo*. Two hours after UV irradiation, binding of F55 to its own promoter was depleted whereas the interaction with that of T_{ind} was greatly weakened (Fig. 6B, panel 2 hours post-irradiation). Subsequently, the promoter regions of the early transcripts T_5 and T_6 were released by F55, indeed, the relative ChIP-sqPCR signals fall to an undetectable level (Fig. 5B, panel 4 hours post-irradiation). Altogether these results indicate that the cascade expression of the SSV1 early genes is regulated through a differential binding of F55 to its target operators.

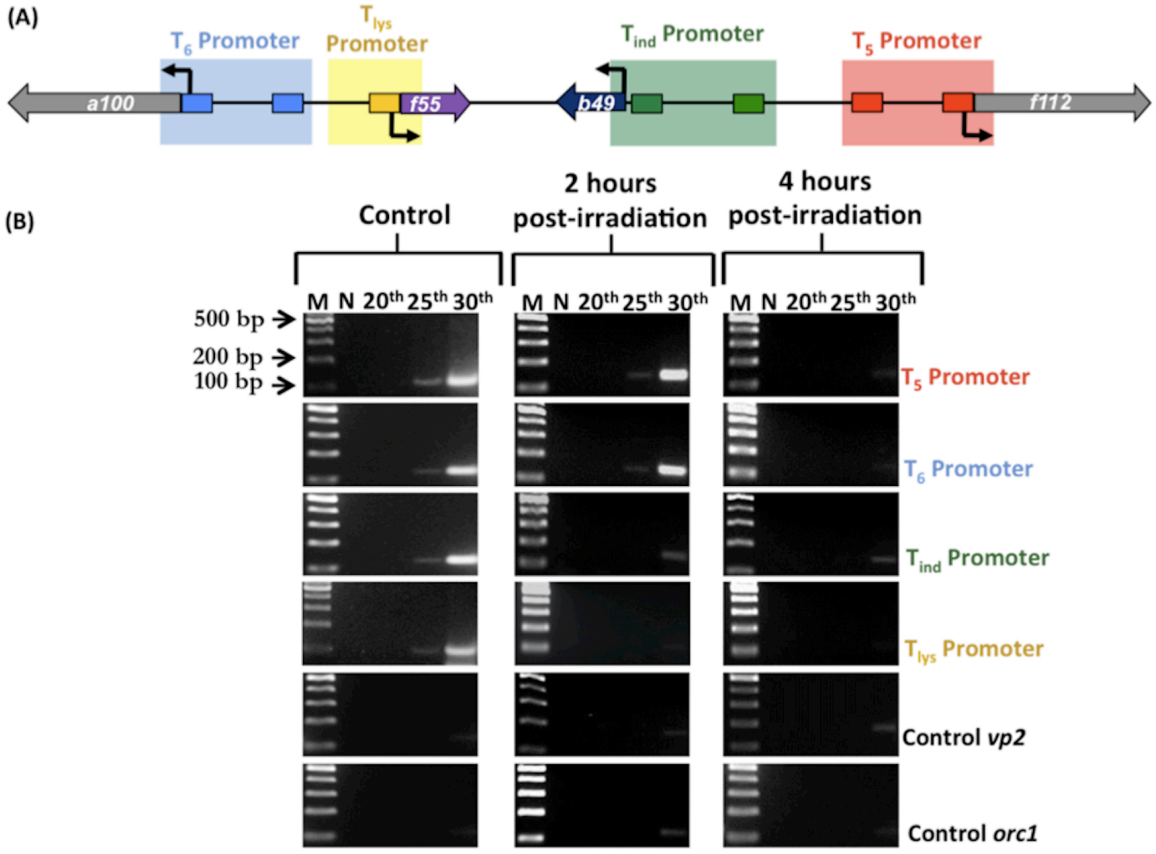


Figure 6. Chromatin immunoprecipitation (ChIPs) on the UV-inducible region of SSV1 using F55 specific antibodies. A) Scheme of the UV-inducible region of SSV1 genome. The amplified regions of the target promoters that are recognized by F55 are blue-, yellow-, green- and red-highlighted. All the promoters contain two F55 binding sites, except for that of T_{lys}. **B)** Total DNA samples prepared from control and UV-treated cultures and collected at 2 and 4 hours post irradiation were used as template for semi-quantitative PCR reactions. Two negative control, *vp2* (viral) and *orc1* (host), were included to assess method specificity. N is the negative control for PCR reactions.

Discussion

Numerous transcription analyses of the UV-inducible fusellovirus SSV1 have been carried out in *S. solfataricus* to dissect viral gene expression patterns before and after UV irradiation. Nevertheless, it was still unclear how this fusellovirus regulates the transition from the lysogeny to the induced status during its life cycle. Our recent work has assigned a key role in this process to the DNA-binding protein F55, which is encoded by the most abundant RNA (T_{lys}) expressed by SSV1 in the carrier stage (23). In order to shed further light on the role of F55 in the regulation of the SSV1 life cycle, an *in vivo* analysis of the events occurring soon after the UV irradiation has been carried out. Several lines of evidence indicate that F55 is indeed the key regulator of the lysogeny/induction transition of SSV1, including: i) F55 is the only transcriptional regulator expressed by SSV1 in the lysogenic status; ii) it binds *in vivo* to the promoters of T_{ind}, T_{lys}, T_5 and T_6 in the absence of UV stimulus, as shown by the ChIP-sqPCR experiments; iii) T_{ind} expression is shut down when F55 is bound to its promoter as shown by Northern blot analysis, and iv) transcription activation of T_{ind} occurs only upon releasing of F55 in the immediate aftermath of the UV irradiation, i.e. within 2 hours. Notably, the transcriptional regulation of T_{ind} is crucial for the lysogeny/induction switch since it is the first RNA whose expression is unlocked in UV-irradiated cells (25).

ChIP-sqPCR data demonstrated that F55 dissociates upon UV irradiation first from the T_{ind} (2 hours post treatment) and subsequently from the T_5 and T_6 promoters (4 hours post treatment) (Fig. 6B). This is in a perfect agreement with its differential affinity towards these regulatory sequences *in vitro* ($T_5 = T_6 > T_{ind}$) (23). Intriguingly, the transcription of T_5 and T_6 is repressed during the SSV1 carrier stage and is activated in UV-irradiated cells at four hours post-treatment (25). These data strongly indicate that F55 regulates the transcriptional activity of these mRNAs in a fashion similar to that of T_{ind}, i.e. by repressing their expression during the lysogenic growth and allowing their transcription activation through clearance of their target sites upon UV irradiation. Although the release of F55 from its targets coincides with the transcriptional activation of the relative genes, it is not possible to exclude the contribution of other factors to enhance their expression levels.

Quantification of the western blot signals demonstrates that the concentration of F55 drops of about 50% in UV-treated cells, if compared to the control sample, and does not dramatically change up to 10 hours post-treatment (Fig. 5). How can these target promoters be progressively released by F55 without fluctuation of the protein levels in the aftermath of UV irradiation? A reasonable hypothesis is that the intracellular concentration of F55 becomes progressively sub-optimal to saturate all the regulative binding sites in the UV-induced cells. Indeed, in the carrier stage, SSV1 constantly replicates in actively dividing cells to keep its copy number around 5 copies per cell (23). In these conditions, F55 saturates 14 sites per SSV1 genome, i.e., 4 sites for each of the T_{ind}, T_5 and T_6 promoters and 2 for that of T_{lys} (Fig. 7), for a total of about 70 binding sites per cell. Upon UV irradiation, whereas the cell division slows down (Fig. 1), the basal viral replication proceeds consistently leading to an increase of the SSV1 copy number starting from the 2nd hour post-treatment (Fig. 2). Thereafter, the episomal SSV1 content reaches ≅160 copies/cell (32) and the total number of sites that has to be bound by F55 rises of about 32 folds. Since in UV-irradiated cells the intracellular level of F55 is constant, the ratio between the number of the binding sites (BS) and the F55 concentration (BS/F55) increases. This variation in the BS/F55 ratio, allows the progressive dissociation of F55 from SSV1 genome that eventually leads to transcription derepression of the target genes (Fig.

6). The proposed mechanism strictly depends on the maintenance of a low and constant F55 concentration in UV-irradiated cells. In this regard, it has been observed in Northern blot experiments that an additional level of control acts post-transcriptionally by degrading T_{lys} (Fig. 4B), probably to compensate the “sloppiness” of the transcriptional control.

Regarding to the transcription regulation of T_{lys}, we did not observe an immediate derepression of its expression although its promoter, likewise that of T_{ind}, is released 2 hours post-irradiation. Indeed, up-regulation of T_{lys} occurs only when T_{ind} transcription is abrogated (8-10 hours post-irradiation, Fig. 4B). It must be taken into account that T_{lys} and T_{ind} RNAs are transcribed in a convergent orientation (Fig. 4A) and therefore the relative genes can be susceptible of transcriptional interference, which is typical of head-to-head oriented gene couples. This architecture strongly resembles that of the non-lambdoid UV-inducible coliphage 186 (35, 36), although a different mechanism may be evoked. Indeed, it has been proved in both eukaryotes (37-39) and prokaryotes (40-42) that transcription of genes distant of a few kilobases is not independent, but coupled by torsional stress that, by influencing the opening properties of the promoters, facilitates or hinders transcription initiation. This kind of influence reduces the expression of convergent gene couples (43), such as in the case of T_{lys} and T_{ind}. This hypothesis seems reasonable in the light of the reported Northern blot and ChIP-sqPCR data. Indeed, the up-regulation of the T_{lys} transcription takes place only after the abrogation of T_{ind} expression.

Altogether these data demonstrate that F55, likewise the CI repressor of lambda, acts as the key switch regulator involved in the lysogeny/induction transition of SSV1. However, some intriguing differences between lambda and SSV1 need to be discussed. Lambda lysogens harbour the viral DNA only as provirus and the only viral gene expressed is the *cI*. The lysogenic/lytic cycle is strictly regulated since viral progeny is produced only upon UV irradiation and causes cell lysis. By contrast, in SSV1 lysogens the provirus co-exists together with some episomal copies and a constitutive extrusion of the viral particles occurs. Therefore, the lysogeny of SSV1 could be better defined as a carrier stage. In this case, the UV irradiation only enhances the rate of SSV1 replication and progeny extrusion without causing cell lysis. This difference is mirrored by the expression regulation of the key regulator F55, which is not as stringent as for the functional homolog CI of lamda phage. From an evolutionary point of view, this kind of host-virus relationship, which is typical for all *Fuselloviridae*, could be co-evolved with the non-lytic nature of these viruses.

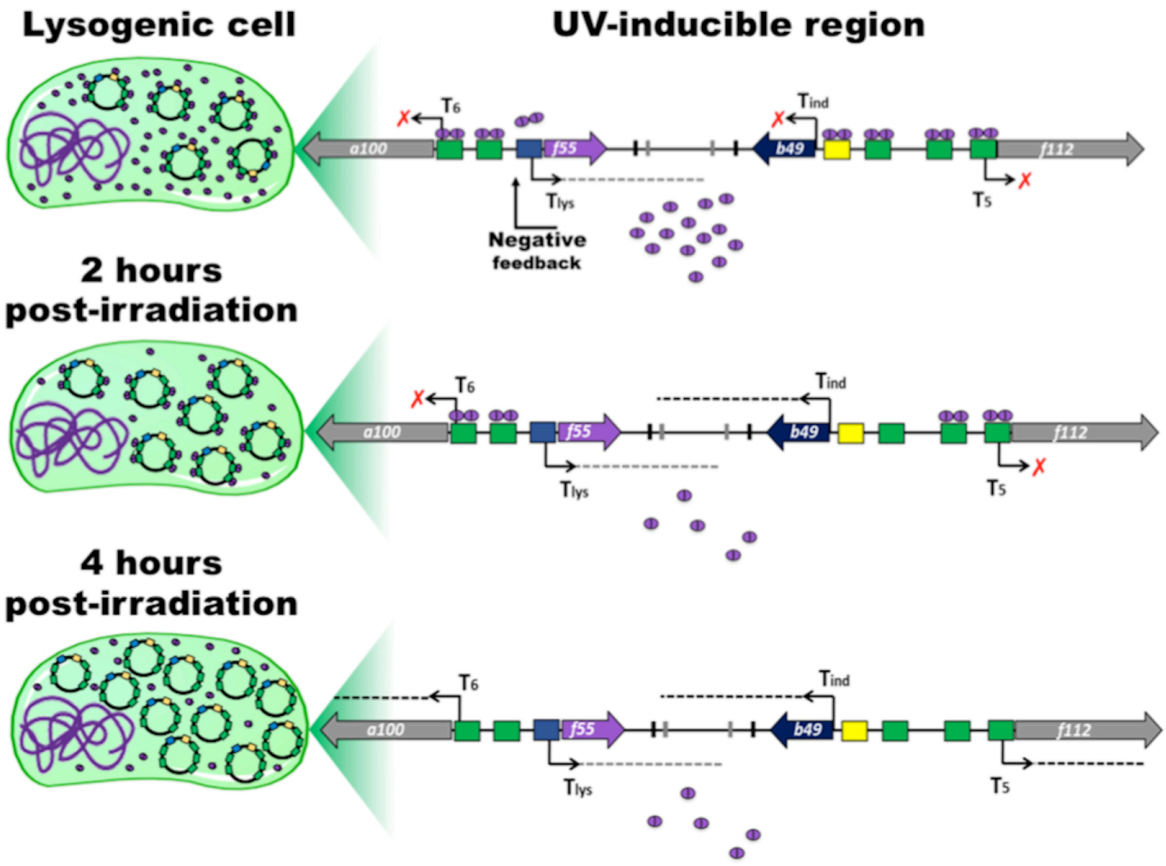


Figure 7. A suboptimal concentration of F55 allows the derepression of the target genes. Schematic representation of the infected cell and of the UV-inducible region of SSV1 genome, are reported. The operators recognized by F55 are green-, yellow- and blue-filled. Black-bent arrows are the transcription start sites and dashed lines represent transcripts. Dimers of F55 are purple-filled ovals. In the **lysogenic cell**, the amount of F55 is suitable to saturate most of its binding sites and to keep SSV1 in a steady carrier stage. **2 hours post-irradiation**, a drop of about 50% of the F55 concentration and a concurrent increase of the viral copy number, lead to the dissociation from the less-affine operators in the promoter of T_{ind} and T_{lys}. Later one, **4 hours post-irradiation**, the dilution effect is enhanced by a further accumulation of the viral DNA, which has as consequence the release of the early promoters (i.e. those of T_5 and T_6), thus allowing transcription derepression.

Acknowledgements

This study was supported by grants from Danish Independent Research Council (DFF-0602-02196B, DFF-1323-00330) and a grant from the Carlsberg Foundation.

References

1. **Prangishvili D.** 2013. The wonderful world of archaeal viruses. Annu Rev Microbiol **8:**565–585.
2. **Contursi P, Fusco S, Cannio R, She Q.** 2014. Molecular biology of fuselloviruses and their satellites. Extremophiles **3:**473–489.
3. **Wang H, Peng N, Shah SA, Huang L, She Q.** 2015. Archaeal extrachromosomal genetic elements. Mol. Micro. Biol. Revs. *In press*, 2015 March
4. **Geslin C, Le Romancer M, Gaillard M, Erauso G, Prieur D**. 2003. Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. Res Microbiol **154:**303–307.
5. **Gorlas A, Koonin EV, Bienvenu N, Prieur D, Geslin C.** 2012. TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus.* Environ Microbiol **14:**503–516.

6. **Sime-Ngando T, Lucas S, Robin A, Tucker KP, Colombet J, Bettarel Y, Desmond E, Gribaldo S, Forterre P, Breitbart M, Prangishvili D.** 2011. Diversity of virus-host systems in hypersaline Lake Retba, Senegal. Environ Microbiol **13:**1956 –1972.
7. **Porter K, Russ BE, Dyall-Smith ML.** 2007. Virus-host interactions in salt lakes. Curr Opin Microbiol **10:**418–424.
8. **Borrel G, Colombet J, Robin A, Lehours AC, Prangishvili D, Sime-Ngando T.** 2012. Unexpected and novel putative viruses in the sediments of a deep-dark permanently anoxic freshwater habitat. ISME J **6:**2119–2127.
9. **López-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcami A.** 2009. High diversity of the viral community from an Antarctic lake. Science **326:**858–861.
10. **Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S, McDermott T, Young MJ.** 2001. Viruses from extreme thermal environments. Proc Natl Acad Sci USA **98:**13341–13345.
11. **Bize A, Peng X, Prokofeva M, MacLellan K, Lucas S, Forterre P, Garrett RA, Bonch-Osmolovskaya EA, Prangishvili D.** 2008. Viruses in acidic geothermal environments of the Kamchatka Peninsula. Res Microbiol **159:**358–366.
12. **Rachel R, Bettstetter M, Hedlund BP, Haring M, Kessler A, Stetter KO, Prangishvili D.** 2002. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. Arch Virol **147:**2419–2429.
13. **Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, Land ML, Ver Berkmoes NC, Hettich RL, Banfield JF.** 2010. Enigmatic, ultrasmall, uncultivated *Archaea*. Proc. Natl. Acad. Sci. U. S. A. **107:**8806–8811.
14. **Comolli LR, Baker BJ, Downing KH, Siegerist CE, Banfield JF.** 2009. Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. ISME J **3:**159–167.
15. **Krupovic M, Quemin ER, Bamford DH, Forterre P, Prangishvili D.** 2013. Unification of the globally distributed spindle-shaped viruses of the *Archaea*. J Virol **88:**2354–2358.
16. **Kraft P, Kümmel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM.** 2004. Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol **78:**7438–7442.
17. **Kraft P, Oeckinghaus A, Kümmel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM.** 2004. Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. J Virol **78:**11544–11550.
18. **Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzer E, Young MJ, Bothner B, Lawrencem CM.** 2008. Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology **376:**270–278.
19. **Guillière F, Peixeiro N, Kessler A, Raynal B, Desnoues N, Keller J, Delepierre M, Prangishvili D, Sezonov G, Guijarro JI.** 2009. Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. J Biol Chem **284:**22222–22237.
20. **Contursi P, D'Ambrosio K, Pirone L, Pedone E, Aucelli T, She Q, De Simone G, Bartolucci S.** 2011. C68 from the *Sulfolobus* islandicus plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. Biochem J **435:**157–166.
21. **Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakić M, Young MJ, Lawrence CM, Copié V.** 2012. Structural studies of E73 from a hyperthermophilic archaeal virus identify the ''RH3'' domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. Biochemistry **51:**2899–2910.
22. **Contursi P, Fusco S, Limauro D, Fiorentino G.** 2013. Host and viral transcriptional regulators in *Sulfolobus*: an overview. Extremophiles. **17:**881–895.
23. **Fusco S, She Q, Bartolucci S, Contursi P.** 2013. T_{lys}, a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. J Virol **87:**5926–5936.
24. **Nadal M, Mirambeau G, Forterre P, Reiter WD, Duguet M.** 1986. Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature **321:**256–258.

25. **Fröls S, Gordon PM, Panlilioc MA, Schleper C, Sensen CW.** 2007. Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology **365:**48–59.
26. **Reiter WD, Palm P, Yeats S, Zillig W.** 1987. Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol Gen Genet **209:**270–275.
27. **Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W.** 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology **185:**242–250.
28. **Schleper C, Kubo K, Zillig W.** 1992. The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sc USA **89:**7645–7649.
29. **Gudbergsdottir S, Deng L, Chen Z, Jensen JVK, Jensen LR, She Q, Garrett RA.** (2011) Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol **79:**35-49
30. **Contursi P, Jensen J, Aucelli T, Rossi M, Bartolucci S, She Q.** 2006. Characterization of the *Sulfolobus* host-SSV2 virus interaction. Extremophiles **10:**615–627.
31. **Wolin EA, Wolin MG, Wolfe RS.** 1963. Formation of methane by bacterial extracts. J Biol Chem **238:**2882–2886.
32. **Fusco S, Aulitto M, Bartolucci S, Contursi P.** 2014. A standardized protocol for the UV induction of *Sulfolobus* spindle-shaped virus 1. Extremophiles. [Epub ahead of print] DOI: 10.1007/s00792-014-0717-y.
33. **Contursi P, Cannio R, Prato S, She Q, Rossi M, Bartolucci S.** 2007. Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. J Bacteriol **17:**6339–6350.
34. **Götz D, Paytubi S, Munro S, Lundgren M, Bernander R, White MF.** 2007. Responses of hyperthermophilic crenarchaea to UV irradiation. Genome Biol **8(10)**:R220.
35. **Callen BP, Shearwin KE Egan JB.** 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol Cell **14:**647–656.
36. **Dodd IB, Egan JB.** 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. Mol Microbiol **45:** 697–710.
37. **Choi CH, Kalosakas G, Rasmussen KØ, Hiromura M, Bishop AR, Usheva A.** 2004. DNA dynamically directs its own transcription initiation. Nucleic Acids Res **32:**1584–1590.
38. **Leblanc BP, Benham CJ, Clark DJ.** 2000. An initiation element in the yeast cup1 promoter is recognized by RNA polymerase ii in the absence of TATA box-binding protein if the DNA is negatively supercoiled. Proc Natl Acad Sci USA **97:**10745–10750.
39. **Alexandrov BS, Gelev V, Yoo SW, Alexandrov LB, Fukuyo Y, Bishop AR, Rasmussen KØ, Usheva A.** 2010. DNA dynamics play a role as a basal transcription factor in the positioning and regulation of gene transcription initiation. Nucleic Acids Res **38:**1790–1795.
40. **Ouafa ZA, Reverchon S, Lautier T, Muskhelishvili G, Nasser W.** 2012. The nucleoid-associated proteins h-ns and fis modulate the DNA supercoiling response of the *pel* genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. Nucleic Acids Res **40:**4306–4319.
41. **Hatfield GW, Benham CJ.** 2002. DNA topology-mediated control of global gene expression in *Escherichia coli.* Annu Rev Gen **36:**175–203.
42. **Du X, Wojtowicz D, Bowers AA, Levens D, Benham CJ, Przytycka TM.** 2013. The genome-wide distribution of non-b DNA motifs is shaped by operon structure and suggests the transcriptional importance of non-b dna structures in *Escherichia coli*. Nucleic Acids Res **41:**5965–5977.
43. **Meyer S, Beslon G.** 2014. Torsion-mediated interaction between adjacent genes. PLoS Comput Biol **10:**e1003785.

2.5 Conclusions

To investigate the relationship between SSV1 and *S. solfataricus* P2, a preliminary whole-transcriptome microarray analysis was carried out in order to study: i) viral gene expression in the carrier stage and ii) the host response to the stable infection of SSV1. As concerns the host, no activation of the CRSIPR-Cas system was observed, thus indicating that this was not challenging the viral infection (further detail are discussed in Chapter III). On the other hand, it was intriguingly the discovery of a new gene on the SSV1 genome, which is actively transcribed and encodes for a putative transcription repressor (F55). Structure prediction of this protein suggested that it could fold into the RHH motif, which is typical of regulator exerting transcription repression of target genes. Functional characterization of F55 revealed that this protein is able to bind *in vitro* to operator sequences located at promoter regions of early-induced genes as well as of its own promoter. Moreover, it has bee proved *in vivo* that the binding of this regulator is associated with transcription repression of the corresponding genes, whereas dissociation of F55 is accompanied by transcription derepression. The above-reported data show how SSV1 controls its gene expression in the carrier stage toward the DNA-binding activity of F55, thus making this virus self-consistent in the regulation of its life cycle in response to external stimuli. Have this ability helped SSV1 to skip the host CRISPR-Cas response? Further evidences will be discussed in the chapter III.

Chapter III

– CRISPR-Cas autoimmunity: when a virus becomes ‘self’–

3.1 Summary

To study host-virus and virus-virus interactions, a whole-transcriptome analysis was carried out on SSV1-, SSV2- and SSV1/SSV2-infected strains of *Sulfolobus solfataricus* P2. As a control, gene expression variations physiologically occurring in an uninfected strain were taken into account. The fuselloviruses SSV1 and SSV2 were chosen because, although they are very similar in morphology, genome organization and gene synteny, their replication induction is induced by different stimuli. In fact, whereas the SSV1 replication is triggered by UV-light exposure (see Chapter II), that of SSV2 is dependent upon the physiological state of the harbouring cell, at least in the natural host *S. islandicus* REY 15/4.

In SSV1-lysogens, the virus expressed few genes including those encoding for structural and structural-related proteins. Moreover, that encoding the transcription repressor F55 was one of the most transcribed. By doing so, SSV1 is able to keep a low copy number by repressing the expression of UV-inducible and early-genes and, at the same time, it produces infecting viral particles throughout the host growth. As concerns the host-counterpart, it was shown that the infection by SSV1 does not significantly affect gene expression of *S. solfataricus*.

On the other hand, it was found that SSV2 gene expression pattern was almost completely overlapping to that of SSV1, except for genes encoding the putative transcription regulator D79 and the DnaA-like protein B223 that is involved in the initiation of the replication. The expression of this latter could be interpreted as an attempt of SSV2 to take the control over the host replication machinery, thus fostering its own replication. Interestingly enough, SSV2 infection caused a huge remodulation of the host expression that involved also the transcriptional activation of almost all CRISPR-Cas loci as well as of several *cas* genes. Even more surprisingly was that the host response was completely reverted by the co-presence of both viruses in the double-infected strain SSV1/SSV2-InF1.

Herein it is shown that the *S. solfataricus* CRISPR-Cas response is consistent with the isolation of single clones showing low SSV2 viral titre and intracellular copy number. Moreover, it is highlighted an interesting aspect of the host-viral interaction in the frame of the CRISPR-response, i.e., host cells containing an integrated provirus are forced to develop a surviving strategy in order to avoid self-attack by the CRISPR-system. Indeed, for the first time, it was shown that *S. solfataricus* cells developed a specific strategy to safeguard host genome integrity, i.e. through deletion of self-targeting spacers. This has cast doubts on the applicability of CRISPR-Cas systems to gain infection-insensitive strains, at least for those viruses that integrate into the host chromosome.

3.2 Transcriptome analysis of *Sulfolobus solfataricus* infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system

Salvatore Fusco[a], Rossana Liguori[b], Danila Limauro[a], Simonetta Bartolucci[a], Qunxin She[c‡#] and Patrizia Contursi[a‡#]

[a]Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, 80126 Napoli, Italy; [b]Dipartimento di Scienze Chimiche, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, 80126 Napoli, Italy; [c]Danish Archaea Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark[b]

[‡]These authors contributed equally to this work.

[#] Address correspondence to: Patrizia Contursi, contursi@unina.it

Qunxin She, qunxin@bio.ku.dk

Keywords: CRISPR-Cas; *Sulfolobus solfataricus*; *Fuselloviridae*; *Sulfolobus* spindle-shaped virus (SSV); Transcriptome analysis.

Submitted to Biochemie

Abstract

The fuselloviruses SSV1 and SSV2 are model systems to investigate on virus-host relationship in stably infected cells thanks to their temperate nature. Although they are very similar in morphology, genome organization and gene synteny, replication induction is regulated by different stimuli, i.e.: by UV-light exposure (for SSV1) and by the physiological status of the host (for SSV2). In this study, we have analysed global gene expression in SSV1- and SSV2-lysogens of *Sulfolobus solfataricus* P2 in the absence of any stimuli Additionally, the interplay among SSV1, SSV2 and *S. solfataricus* has been investigated in a double-infected strain to explore both virus-host and virus-virus interactions. Whereas SSV1 did not induce major changes in the host gene expression, SSV2 elicited a strong host response, which includes the transcriptional activation of CRISPR loci and *cas* genes. As a consequence, a significant decrease of the SSV2 copy number has been observed, which in turn led to provirus-capture into the host chromosome. This study has revealed novel aspect of the host-viral interaction in the frame of the CRISPR-response that is herein presented.

Introduction

The majority of organisms are susceptible to viral infection and accordingly it has been predicted that more than a thousand of different viruses inhabit several niches worldwide (Suttle 2007; Comeau et al., 2008). This makes viruses, especially those infecting prokaryotes (i.e. *Archaea* and *Bacteria*), the most predominant biological entities on Earth (Pietilä et al., 2014). Soon after the discovery of the *Archaea* domain, many viruses and virus-like particles have been isolated from extremely hot, low pH or hypersaline niches (Prangishvili 2013; Wang et al., 2015). Intriguingly, these viruses exhibit unique morphologies comparing to those infecting *Bacteria* (another domain of prokaryotes) and fall into 10 new virus families in the current virus classification (Pina et al., 2011; Peng et al., 2012; Prangishvili 2013). For instance, spindle-shaped viruses are exceptional on their own since this morphotype is a hallmark of viruses infecting *Archaea* (Krupovic et al., 2013) and it has never been described for bacteriophages or eukaryal viruses.

Spindle-shaped viruses infecting organisms of the genus *Sulfolobus* belong to the family *Fuselloviridae*, which comprises so far ten members (SSV1, SSV2, SSV3, SSV4, SSV5, SSV6, SSV7, SSV8, SSV9 and ASV1) (Contursi et al., 2014a; Wang et al., 2015). *Sulfolobus* spindle-shaped virus 1 (SSV1) is a model virus for investigating virus-host relationship in *Archaea*, since it is the only UV-inducible archaeal virus isolated so far (Fröls et al., 2007; Contursi 2014a; Fuso et al., 2014). Investigation of SSV1 transcription has laid the basis for understanding how gene expression is regulated in *Archaea* (Reiter et al., 1988a; 1988b). Moreover, its genome has served as a backbone for the construction of vectors for genetic manipulation and gene expression in *Sulfolobus* (Contursi et al., 2013). SSV1 is a temperate virus that, upon infection, establishes a stable coexistence with the host by keeping its copy number low and constant throughout the growth of infected cells. It is worth noting that, unlike lambda-lysogens where only provirus exists, SSV1-lysogens carry both a provirus (an integrated viral genome in the host chromosome) and a few episomal copies of the viral DNA (Nadal et al., 1986; Schleper et al., 1992). Consequently, viral progeny is constitutively produced at a low level even in the absence of the inducing stimulus (i.e. UV-light exposure). Therefore, the lysogenic state of SSV1 is better defined as a carrier stage (Fusco et al., 2013 and 2015 submitted). The transcriptional map of SSV1 highlighted the chronological expression of the viral genes as consequence of

the UV exposure and helped speculating about the function of their products. However, so far, the general lack of sequence similarity with other proteins in public database has hindered functional studies of these viral genes. Indeed, only a limited number of proteins encoded by SSV1 or other fuselloviruses have been functionally and/or structurally characterized (Kraft et al., 2004a; 2004b; Menon et al., 2008; Guillière et al., 2009; Contursi et al., 2011; Schlenker et al., 2012; Contursi et al., 2013; Contursi et al., 2014b; Fusco et al., 2013).

On the other hand, although SSV2 resemble SSV1 in the shape of the viral particle as well as in genome organization and gene synteny, its replication induction is regulated by the physiological state of the host cell rather than by an external stimulus as demonstrated for SSV1. After infecting the native host *S. isandicus* REY15/4, the copy number of the virus is kept low (around 1-3 copies per cell) until the host enters into the stationary growth phase in which a steep increase of the copy number occurs (25-50 copies per cell) (Contursi et al., 2006). Noteworthy, this effect has not been observed for the permissive host *S. solfataricus*, in which SSV2 shows a copy number comparable to that of the induced state of the natural host throughout the growth of *S. solfataricus*. This led to hypothesis that a transcription factor encoded by the natural host *S. islandicus* was likely involved in the regulation of the viral replication induction and this factor is presumably absent from the foreign host *S. solfataricus* (Contursi et al., 2006).

Recently, a transcriptome analysis has been carried out to define the gene expression pattern of SSV2 upon infection of *S. solfataricus* cells (Ren et al., 2013). This allowed detecting seven non-overlapping transcripts that were termed after the SSV1 counterparts. Moreover, genome transcription occurred, as for SSV1, in a temporal fashion during the very early phase of the infection. Nevertheless, early genes were not adjacently located and displayed a distributive pattern of expression (Ren et al., 2013). The analysis of the SSV2 gene expression pattern in the carrier stage (i.e. once the virus has settle down into the host cell) would allow identifying a minimal set of genes required for the maintenance of a basal virus-host interaction. Nevertheless, this aspect has received little attention so far.

Only a few global gene expression analyses of hosts infected with archaeal viruses have been reported (Fröls et al., 2007; Ortmann et al., 2008; Okutan et al., 2013; Ren et al., 2013; Quax et al., 2013). Worth-mentioning are those performed on the lytic *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2), which exhibits a chronologically coordinated gene expression (Okutan et al., 2013; Quax et al., 2013) as well as on the *Sulfolobus* turreted icosahedral virus (STIV), which does not show any temporal regulation of genes expression (Ortmann et al., 2008). Viral-host interactions in *Crenarchaea* have been successful investigated at gene expression level both for lytic and temperate viruses, using as host *S. solfataricus* P2 isolated from a hot spring in Naples (Italy). *S. solfataricus* is a suitable crenarchaeal model to study fusellovirus-host relationships, indeed, it has been shown that several strains isolated in Italy are susceptible to infection by all fuselloviruses isolated to date (Ceballos et al., 2012). Moreover, the strain P2 of *S. solfataricus* harbours the CRISPR-Cas antiviral defense system, which is one of the main players of the virus-host arms race. A CRISPR locus is an array of repeats, which are identical in sequence and length within the same locus and are interspaced by spacer sequences. In brief, to achieve viral immunity, DNA sequences of invading genetic elements (i.e. protospacers) are integrated in the array of a CRISPR locus in a process known as “adaptation”, thus becoming new spacers of the CRISPR loci. Transcription of these loci produces long-crRNAs (pre-CRISPR RNAs), which are

processed to produce mature crRNAs. Subsequently, mature crRNAs form ribonucleoprotein complexes with Cas proteins to detect and degrade the foreign nucleic acid in the stage known as “interference”.

The genome of *S. solfataricus* P2 harbors six CRISPR loci (from A to F) and cas gene cassettes encoding for up to seven different interference complexes, i.e. of subtypes I-A (Cascade complex), III-B (CMR complex) and III-A (Zhang and White 2013). Whereas ribonucleoprotein complexes of subtypes IA and III-A excerpt DNA recognition and degradation, those belonging to the subtype III-B cleave RNA in a sequence-specific manner. These features make *S. solfataricus* P2 an exceptional model to investigate the modulation of the CRISPR-Cas system activity in response to the infection by different viruses. In this regard, we report herein a global gene expression analysis that highlights gene expression remodulation in *S. solfataricus* upon the stable establishment of SSV1 and SSV2 into the host cells. Previous studies have been carried out to investigate on variation of gene expression in the immediate aftermath of SSV2 infection (up to 10 hours later) or upon UV-stimulus in SSV1 lysogens. We show that once SSV1 and SSV2 settle down into the cells, the host response elicited by the two viruses is very different. Noteworthy, the up-regulation of CRISPR system occurs only in cells infected with SSV2.

Materials and methods

Strains, media and growth conditions

SSV1-, SSV2- and SSV1/SSV2-infected strains of *Sulfolobus solfataricus* P2 were generated, as described elsewhere (Contursi et al. 2006, Fusco et al. 2013), using as host the uracil auxotrophic mutant InF1 (Gudbergsdottir et al. 2011). Cultures were grown aerobically in TYSU, i.e. a glycine-buffered Brock’s basal salt solution supplemented with 0.1% tryptone, 0.05% yeast extract, 0.2% sucrose and 0.002 % uracil (w/v), being the final pH 3.2 upon addition of concentrated H_2SO_4. Incubation was conducted in 250-ml Erlenmeyer flasks at 75 °C, with a shaking rate of 150 rpm in an Innova 3100 Water bath shaker (New Brunswick Scientific Corp).

Aliquots from frozen cultures of the uninfected InF1 and the infected strains (SSV1-InF1, SSV2-InF1 and SSV1/SSV2-InF1) were revitalized by inoculating in TYSU medium. Cell growth was spectrophotometrically monitored at 600 nm (OD_{600}) throughout the cultivation by means of a Variant Cary® 50 Bio UV/Visible Spectrophotometer (McKinley Scientific). Once reached 0.4 OD_{600}, cultures were diluted to 0.05 OD_{600}, incubated back to 75°C and samples collected at 0.4 OD_{600} (early exponential phase) and 1.2 OD_{600} (late exponential phase). Cellular pellets were obtained through centrifugation at 3,000 × *g* for 10 min using the Centrifuge 5810R (Eppendorf) and treated for total DNA and RNA preparations.

To isolate single clones from the SSV2-InF1 strain, serial dilutions of this culture were plated on TYSU-Gelrite and incubated at 75°C. Isolated colonies appeared on the plate surface after 7-10 days (about 100 colonies per plate). Several colonies were inoculated in liquid medium (about 10 per plate), let to grow until 0.4 OD_{600} and culture supernatant tested by plaque assay for viral titre determination, using the uninfected InF1 strain as lawn and the supernatant of the SSV2-InF1 culture as control. Single clones showing a lower viral titre were further screened through three subsequent steps in order to select SSV2-cured cells.

RNA extraction, cDNA synthesis and labeling

Total RNA samples were prepared by means of the TRIzol reagent (Sigma Aldrich®) and carried-over DNA was digested using Turbo™ DNase (Ambion®),

according to the manufacturer's instructions and the enzyme was thermal inactivated at 70°C for 10 min after the addition of 5 mM EDTA. DNA-free RNA samples were purified by phenolic extraction and ethanol precipitation. RNA pellets were dissolved in nuclease-free water and both concentration and integrity were checked through: i) electrophoresis on denaturing, formaldehyde-containing 2.0% (wt/vol) agarose gel and ii) determination of the 260nm/280nm adsorption ratio using a Nanodrop 1000 Spectrophotometer (Thermo Scientific). Only samples showing ratios between 2.1 and 1.9 were used for cDNA synthesis.

cDNA labeled with 5-(3-aminoallyl)-dUTP (aa-dUTP) was generated by means of the Amersham CyScribe Post Labeling Kit (GE Healthcare) according to the manufacturer's directions, with few modifications as follows: 4 µl of random hexamers primers, 1 µl random nonamers primers, 10-15 µg total RNA and nuclease-free water were mixed in a final volume of 11 µl. The mixture was denatured at 70°C for 5 minutes and chilled at room temperature for 10 minutes. 4 µl of 5× CyScribe buffer, 1 µl dNTP mix, 1 µl aa-dUTP, 2 µl DTT 0.1 M and 1 µl CyScribe reverse transcriptase, were added to the reaction mixture and incubated at 42°C for 90 min. Afterwards, the RNA template was degraded by adding 2 µl of 2.5 M NaOH and incubating at 37 °C for 15 minutes. The allyl-dUTP cDNA sample was purified using the Illustra Cyscribe GFX Purification kit (GE Healthcare) after neutralizing the reaction mixture with the addition of 10 µl of 2 M HEPES free acid. The concentration of the purified cDNA was spectrophotometrically determined.

The CyDye labeling of the amino allyl-modified cDNA was achieved using the Amersham CyScribe Post Labeling Kit (GE Healthcare) and purification of the labelled cDNA was carried out by means of the MinElute® PCR purification kit (Qiagen), following the manufacturer's instructions. The "reference" and the "experiment" cDNA samples were labeled with cyanine-3 (Cy-3) and cyanine-5 (Cy-5), respectively. The ratio between the CyDye-labeled and the total cDNAs was used to monitor the labelling efficiency. Only "reference" and the "experiment" cDNA samples showing similar labelling efficiency were co-hybridized on the same slide.

Microarray hybridization and data analysis

Customized microarray slides harbor about of 7,000 spots and were designed by the *Sulfolobus* genome chips consortium and manufactured by Ocimum Biosolutions (Hyderabad). A single array includes probes, spotted in duplicate, for 3,042 *S. solfataricus* P2 genes, for several crenarchaeal viruses, three plasmids as well as for human and *Arabidopsis thaliana* sequences as negative controls (Ortmann et al., 2008, Okutan et al., 2013, Ren et al., 2013). The microarray slide was first dipped into a coupling jar containing 25 ml of prewarmed prehybridization solution (1% BSA; 5× SSC; 0.1% SDS) and then incubated at 42°C for 40 minutes under shaking (Thermo-Electron Corporation). The prehybridized slide was washed three times with distilled water and once with 100% isopropanol. Finally, the slide was dried by centrifugation and a LifterSlip coverslip was applied onto the array-area.

The hybridization solution (60% deionized formamide, 7× SSC, 2% SDS, 0.2 µg/µl herring sperm DNA, 0.2 µg/µl tRNA) was incubated at 95°C for 2 minutes and chilled on ice for 1 minute, before adding equal amounts of reference and the experiment cDNA (50 picomoles CyDye for each). Subsequently, the hybridization mixture was carefully injected between the array surface and the coverslip. The slide was sealed in a hybridization chamber and incubated at 42°C for 18 hours. Washing of the hybridized slide was performed with: i) a prewarmed solution A (2× SSC; 0.1% SDS) at 42°C for 5 minutes under gently shaking; ii) a prewarmed solution B (0.1×

SSC; 0.1% SDS) for 20 minutes at 42°C and iii) a solution C (0.1× SSC) for five times at room temperature. Finally, the slide was dried by centrifugation and immediately scanned using the Array Wor$_X^e$ (Applied Precision). Data analysis was conducted by ImaGene® v. 9.0 (BioDiscovery), using default settings and included the following steps: data import, background adjustment, normalization, summarization and quality assessment.

Up to four independent experiments were carried out for each reference/experiment cDNA couple. Moreover, since each probe is present in duplicate on a slide, the fold-change variation for each gene is the average among the collected data points. In particular, the software Imagene® gives a $\log_2$ ratio value for each analysed spot:

$$[\log_2 (\text{FCy-5 exp})] – [\log_2 (\text{Fcy-3 ref})] = \log_2 (\text{Fcy5 exp/ Fcy3 ref})$$

where “F” indicates the normalized fluorescence intensities of the Cy-3 and Cy-5 detected in a given spot. The variation of the gene expression fold (VGEF) is calculated through the following equation:

$$\text{VGEF} = 2^{\log2(\text{Fcy-5exp/ Fcy-3ref})}$$

Only genes with a VGEF of ≥ 2 as well as a *p* value of < 0.05 were regarded as differentially expressed.

PCR analysis of CRISPR loci leader-proximal regions and of the SSV2 integration site

A fundamental step in the CRISPR-Cas system is the *adaptation*, which consist in the Cas proteins-mediate incorporation of new spacers that occurs immediately downstream the leader sequence of a given CRISPR locus. With the purpose of detecting new spacers integration at the leader-proximal regions of all the CRISPR loci (A-F) in the *S. solfataricus* genome, six primer couple were used (Tab. 1). In particular, these oligonucleotides were designed to amplify the first 5-8 repeat-spacer units located immediately downstream the leader sequence (Erdman and Garrett 2012). Thus, integration of new spacers will produce longer PCR products than the control sample (uninfected InF1).

Total DNA samples were prepared using the DNeasy tissue kit (Qiagen), following the manufacturer’s instructions and their concentration was spectrophotometrically measured using a Nanodrop 2000 Spectrophotometer (Thermo-Scientific). PCR master mixes were prepared as follows: 1× Taq Buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP mix, 0.6 µM *primer-fw*, 0.6 µM *primer-rv* and 0.05U/µl of Taq DNA Polymerase (Thermo Scientific). The thermal cycling protocol was as follows: an initial denaturation step of 5 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 50°C, and 1 min at 72°C. A final step at 72°C has been carried out for 10 min at the end of the 35th cycle. A negative (no template) was also included in the analysis. PCR products were analysed in a 1% agarose gel.

In parallel, to check the occupancy of the SSV2-integration site in the *S. solfataricus* genome, a PCR analysis was carried out by means of the primers SSV2attAp01, SSV2attAp02 and SSV2attAp03 (described in Contursi et al. 2006), whose sequences are listed in Table 1. The reactions were set up (as described above) except for the annealing temperature that was of 55°C.

Table 1
Primers used for PCR analyses

Name	Sequence (5'-3')	Length (nt)
Lpr-LcsA-fw	AGCTTCTGACCCGCTCCTGA	20
Lpr-LcsA-rv	GCACATCATCAAACAATGGTAAGCC	25
Lpr-LcsB-fw	AGGGGTTTGTGGGATGGGTTGTG	23
Lpr-LcsB-rv	ACAACTACCACCACTACCACGG	22
Lpr-LcsC-fw	TCGCTTATCTCTCTCATGCGCCATT	25
Lpr-LcsC-rv	TGTCCCGTTTTTGTAAGTGGGGG	21
Lpr-LcsD-fw	AGTTCCACCCCCGAAGCTCCT	21
Lpr-LcsD-rv	AGCCGGGACAAGTTTCACAAATTGA	25
Lpr-LcsE-fw	ATAGGGAAAGAGTTCCCCCG	20
Lpr-LcsE-rv	TGACTCTAGTGCAATCTTCGA	21
Lpr-LcsF-fw	CGGCGTTATAATGGGTATCGGAATCGG	27
Lpr-LcsF-rv	GCTCACTATCTCACCCCTATCAATACCC	28
SSV2attAp01	GTGTTCTACCTTTTCCACAGTC	22
SSV2attAp02	TGGGTACGTCATTTATTGATCTT	23
SSV2attAp03	GCTTTTATGCAGTTATTGCTTT	22
orc1-fw	TATAAATTGTTATAGACATAGAACGCTGTA	30
orc1-rv	TTAAATACTTCTTGTGCCGATAGTCC	26
vp3-fw	GTTATTGGTGTAGTTCTGTT	20
vp3-rv	GGGTACAACAATTAAGACTA	20

Semi-quantitative PCR analysis of the SSV2 DNA content

Cell pellets of the initial population InF1-SSV2 and of the screened clones showing a lower SSV2 titre (clone 1, 1Q and $1Q_1$) were treated for total DNA extraction using the DNeasy tissue kit (Qiagen), following the manufacturer's instructions. To follow the variation of the SSV2 DNA content among the different clones, a semi-quantitative PCR analysis was performed. Two primer couples were designed using Primer3 software (available at the website: http://bioinfo.ut.ee/primer 3-0.4.0/), in order to amplify: (i) a 108-bp region of the host single-copy gene *orc1* and (ii) a 238-bp region of the SSV2 single-copy gene *vp3* (Table 1). A master mix was prepared and PCR reactions were carried out as describe elsewhere (Fusco et al., 2014), using the following cycling protocol: an initial denaturation step of 5 min at 95°C, followed by 30 cycles of 40 sec at 95°C, 40 sec at 62°C, and 1 min at 72°C. For each reaction, tubes were taken from the thermocycler at the 20^{th}, 25^{th} and 30^{th} cycle of amplification. PCR products were run a 2% agarose gel in 1× TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA).

Results and discussion

Two fuselloviruses, SSV1 and SSV2, were chosen as model study since they are not lytic and are therefore suitable for analysing variation of gene expression in stable infected populations, i.e. once the host/virus relationship has been settled down (carrier stage). Indeed, previous studies were carried out in the immediate aftermath of the viral infection (Ren et al., 2012), on lytic viruses (Ortmann et al., 2008, Okutan et al., 2013, Quax et al., 2013) or upon induction of the viral replication (Fröls et al., 2007). Furthermore, the double-infected SSV1/SSV2-InF1 strain is a ternary system, which is suitable for analysing the mutual effect of the two viruses on each other as well as on the host gene expression.

Viral gene expression in the infected strains

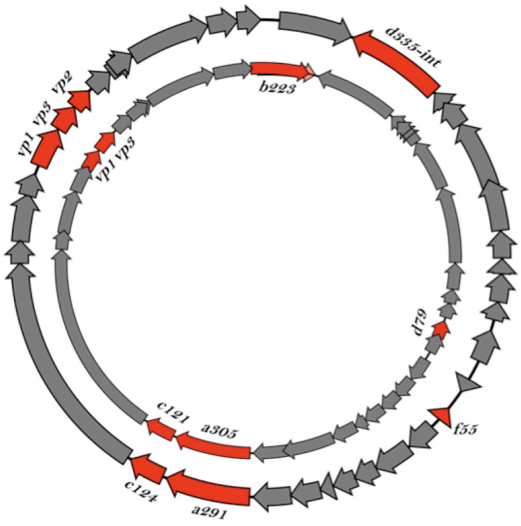
To evaluate the effect of the viral infection on the host gene expression, *S. solfataricus* P2 (InF1) was infected to generate virus harbouring stains with a single virus, including SSV1-InF1, SSV2-InF1 and SSV1/SSV2-InF1; this latter carries both fuselloviruses. SSV1 and SSV2 genes show sequence identity and genome synteny and, accordingly, SSV2 promoters were named after their SSV1 counterparts (Ren et al., 2013; Contursi et al 2014). The only significant difference is the absence, in the SSV2 genome, of a region similar to that of SSV1 that is involved in regulating the switch from the lysogenic to the UV-induced state. Microarray experiments were performed for all the strains, at two different phases, i.e. at 0.4 OD_{600nm} (early-exponential) and 1.2 OD_{600nm} (late-exponential), in order to define which viral genes were expressed. With this aim, microarray slides were co-hybridized with the total cDNA samples from:

1) uninfected InF1 vs infected SSV1-InF1, collected at 0.4 OD_{600nm}
2) uninfected InF1 vs infected SSV1-InF1, collected at 1.2 OD_{600nm}
3) uninfected InF1 vs infected SSV2-InF1, collected at 0.4 OD_{600nm}
4) uninfected InF1 vs infected SSV2-InF1, collected at 1.2 OD_{600nm}
5) uninfected InF1 vs infected SSV1/SSV2-InF1, collected at 1.2 OD_{600nm}
6) uninfected InF1 vs infected SSV1/SSV2-InF1, collected at 1.2 OD_{600nm}

cDNA were prepared from the uninfected and infected strains and used as reference and experimental samples, respectively. These comparisons allowed us to identify viral genes expressed in the infected strains. SSV1 genes found to be expressed at both the growth phases analysed were those encoding: i) the structural proteins VP1, VP3 and VP2, ii) the integrase D335, iii) the transcription repressors F55 and iv) A291 and C124 (Fig. 1) for which virion docking/release and structural functions have been proposed, respectively (Fröls et al., 2007, Ren et al., 2013). Interestingly, most of the viral gene expressed by the SSV2-InF1 strain were homologous to those expressed from SSV1, i.e.: VP1, VP3, A305 (SSV1-A291) and C121 (SSV1-C124) (Fig. 1). In addition, the DnaA-like protein B233 and the D79 were found expressed. A probe matching the 3' non-coding region of *b233* gene detected transcriptional activity, thus indicating that the mRNA of *b233* extends over about a hundred nucleotides downstream its own stop codon (Fig. 1). In the double-infected strain (SSV1/SSV2-InF1), viral genes expressed were those encoding: i) the SSV1 and SSV2 structural proteins VP1 and VP3, ii) the homologous proteins A291/A305 and iii) C124/C121, iii) the transcription repressor F55 of SSV1 and iv) B233 and D79 of SSV2 (Fig. 1). It is noteworthy that the transcription patterns of SSV1 and SSV2 were nearly identical, despite the fact that replication induction for these two viruses is

triggered by different stimuli, i.e. the UV-light exposure and the host physiological/metabolic state, respectively. Therefore, this analysis defined a minimal set of genes required, in the carrier stage, for the replication and packaging of both fuselloviruses (Fig. 1).

(A)



(B)

Viral genes expressed				
SSV1-InF1	SSV2-InF1	SSV1/SSV2-InF1		Function
SSV1-vp1	*SSV2-vp1*	*SSV1-vp1*	*SSV2-vp1*	Structural
SSV1-vp2	[a]H.N.D.	[c]N.E.	[a]H.N.D.	Structural
SSV1-vp3	*SSV2-vp3*	*SSV1-vp3*	*SSV2-vp3*	Structural
d335	[b]H.N.E.	[c]N.E.	[c]N.E.	Integrase
f55	[a]H.N.D.	*f55*	[a]H.N.D.	Transcription repressor
a291	*a305*	*a291*	*a305*	Putative viral docking/release
c124	*c121*	*c124*	*c121*	Putative structural
[b]H.N.E.	*b233*	[b]H.N.E.	*b233*	DnaA-like protein
[b]H.N.E.	*d79*	[b]H.N.E.	*d79*	Unknown

[a] Homologue not detected, [b] Homologue not expressed, [c] No expression

Figure 1. Schematic representation of the SSV1 and SSV2 genomes. (A) Red-filled arrows are viral genes expressed in the carrier stage for both the SSV1 (outer) and the SSV2 (inner) genomes. Clockwise- and anticlockwise-oriented arrows represents ORFs encoded by the plus and minus strand, respectively. (B) List of viral gene expressed in all the stains analyzed.

In order to evaluate variation of the expression levels of the viral genes during the host growth, the following co-hybridization were carried out:

1) SSV1-InF1 collected at 0.4 OD_{600nm} vs SSV1-InF1 collected at 1.2 OD_{600nm}
2) SSV2-InF1 collected at 0.4 OD_{600nm} vs SSV2-InF1 collected at 1.2 OD_{600nm}
3) SSV1/SSV2-InF1 collected at 0.4 OD_{600nm} vs SSV1/SSV2 collected at 1.2 OD_{600nm}

cDNA from the infected strain collected at 0.4 OD_{600nm} and 1.2 OD_{600nm} were used as reference and experimental samples, respectively. This analysis revealed a constitutive expression of all the above-listed SSV1 and SSV2 genes, except for the SSV2 ORF *a305*, which was down-regulated of 2.9-fold (Table S3 and S4).

Expression variation in the uninfected InF1 strain throughout the growth

To evaluate the effect of viral infection on the host gene expression, we first identified the up- and down-regulated genes during the growth of the uninfected InF1 strain. With this aim, microarray slides were co-hybridized with total cDNA samples from:

1) InF1 collected at 0.4 $OD_{600nm(ref)}$ vs InF1 collected at 1.2 $OD_{600nm(exp)}$

A total of 60 up-regulated and 27 down-regulated genes were detected (Tab. 2). Among the up-regulated ones, the expression of 38 genes increased of 2.0-4.9 folds, 13 showed a 5.0-9.9 fold-change and only for 5 genes the expression was up-regulated among 10.0 and 33.0 folds. On the other hand, all down-regulated genes showed a fold of repression comprised in the range 2.0-4.9, except for one (*sso2574*) whose transcription was down-regulated by 5.9 folds.

A remarkable number of the up-regulated *S. solfataricus* genes encode proteins belonging to three functional categories: *hypothetical protein*, *transport* and *energy metabolism* (Tab. 2). The transcriptional activation of genes involved in the transport of small metabolites is not surprising and it is likely due to the shortage of nutrients, which occurs in a later stage of the growth. Moreover, it is interesting that a homologous of the gene encoding the bacitracin resistance protein (*sso1860*) is up-regulated, hinting to defence mechanisms acting against xenobiotic peptides produced by other member cells to face overpopulation. On the other hand, down-regulated genes mostly cluster into the categories of *transcription and regulation* as well as *translation* (Tab. 2). Noteworthy, the down-regulation of many genes encoding for ribosomal proteins is probably consistent with the reduction of the proteins synthesis and, in turn, with the decrease of cell growth rate. A complete list of the differentially regulated genes, grouped by functional categories, is reported in the Table S1.

Table 2
Functional distribution of up- and down-regulated genes.

Functional Group[a]	Gene expression							
	Up-regulated genes				Down-regulated genes			
	InF1	SSV1 InF1	SSV2 InF1	SSV1/SSV2 InF1	InF1	SSV1 InF1	SSV2 InF1	SSV1/SSV2 InF1
Aminoacid Biosynthesis	2	1	2	1	2	0	2	0
Cellular Envelope and Membrane	1	0	1	0	0	0	1	0
Cellular Process	2	1	4	2	2	1	8	2
Central Intermediary Metabolism	0	2	0	2	0	0	2	0
Cofactor Biosynthesis	0	0	0	0	0	0	2	1
Energy Metabolism	10	1	19	0	1	0	12	3
Hypothetical Protein[b]	16	3	10	1	6	4	16	3
IS Elements	1	2	6	0	1	0	0	0
Lipid Metabolism	2	0	3	1	0	0	0	0
Nucleotides	1	0	0	0	0	0	7	1
Protease and Protein Modification	0	0	0	0	1	0	2	0
Replication and Repair	1	0	1	0	0	0	3	2
RNA	2	1	1	1	0	1	0	0
Transcription and Regulation	3	1	11	2	5	2	7	2
Translation	0	0	0	0	7	4	14	0
Transport	17	4	7	1	1	1	5	2
Uncategorized Helicases	1	1	3	0	1	0	2	2
Cas genes	0	0	8	0	0	0	0	0
Viral Genes	0	0	0	0	0	0	1	1
STIV infection differentially regulated	1	0	1	0	0	0	1	0
Total	60	17	77	11	27	13	85	19

[a]Putative functions are derived from the website of the *Sulfolobus solfataricus* P2 complete sequencing project. [b]Hypothetical proteins with no putative function.

Host gene expression variations as consequence of SSV1 infection

In order to detect remodulation of the host gene expression in the SSV1-InF1 lysogenic strain, a microarray analyses was carried out by co-hybridizing cDNA samples as follows:

1) SSV1-InF1 collected at 0.4 $OD_{600nm(ref)}$ vs SSV1-InF1 collected at 1.2 $OD_{600nm(exp)}$

To determine the host genes whose expression is altered as a consequence of the SSV1 infection, we excluded from the analysis those that were up- and down-regulated in the uninfected InF1 strain as a consequence of the physiological changes occurring in the passage from the early to the late exponential growth phase. Among the thirty differentially expressed host genes, 17 were up- and 13 down-regulated throughout the growth of the SSV1-InF1 strain (Tab. 2). Changes in the expression level tended to be higher for the up-regulated genes. The majority of the down-regulated ones decreased by 2 to 2.5-folds, whereas 5 genes were up-regulated with a increasing from 2.8 to 4.3-fold (Tab. S2). Although these genes are scattered over several functional categories, a similar trend to that of the uninfected strain was observed. Indeed, genes encoding proteins involved in the *transport* and *translation* are up- and down-regulated, respectively.

These data show that the infection by SSV1 has only a limited effect on the host gene expression, suggesting a harmonic coexistence with the host in the carrier stage. Accordingly, SSV1 is a temperate virus that self-regulates its gene expression in the carrier stage through the pleiotropic effect of the transcription regulator F55 (Fusco et al., 2015 submitted), which represses the expression of the UV-inducible as well as of the early genes.

Re-modulation of the host gene expression by SSV2 infection: the case of the CRISPR-Cas system activation

To investigate on variation of host gene expression during the growth of the stably infected SSV2-InF1 strain, a microarray analysis was carried out by co-hybridizing cDNA samples as follows:

SSV2-InF1 collected at 0.4 $OD_{600nm(ref)}$ vs SSV2-InF1 collected at 1.2 $OD_{600nm(exp)}$

As already described for the SSV1-InF1 strain, all the variations of gene expression reported herein for the SSV2-InF1 strain, have to be considered merely due to the SSV2 infection. Noteworthy, a total of 162 host genes were differentially expressed throughout the growth of the SSV2-InF1 strain, with 77 and 85 genes that were found to be up- and down-regulated, respectively (Tab. 2). By comparing the total number of differentially regulated host genes between the SSV1-InF1 (30 genes) and the SSV2-InF1 (162 genes) strains, it is evident that the remodulation occurring as consequence of the SSV2 infection is outstanding (Tab. 2). Expression variation is in general higher for up-regulated genes (between 2.0 and 7.9 folds) than for those down-regulated (between 2.0 and 6.0 folds). Moreover, the expression of the majority of down-regulated genes decreased by 3.0-folds or less, whereas 34 up-regulated genes were differentially regulated of 4-folds or greater (Tab. S3). Intriguingly, although SSV2 is a not-lytic virus, its effect on the host gene expression is comparable to the lytic virus STIV (Ortman et al., 2008), indeed, 5% of the *S. solfataticus* genes were directly affected by the SSV2 infection.

Among these, genes encoding for proteins belonging to the functional categories of energy metabolism as well as transcription and regulation tended to be up-regulated (Tab. 2). Furthermore, induction of IS elements transcription (Tab. S3) has already been reported after viral infection for the lytic virus SIRV2 (Quax et al 2013) as well as consequently to UV-light (Fröls et al., 2007) and heat shock exposure (Cooper et al., 2009), thus highlighting their involvement in the general stress response in *S. solfataricus*. On the other hand, down-regulated genes are mostly associated with categories of translation (Tab. 2 and S3). Moreover, repression of stress response genes encoding for the proteasome subunit (*sso0278*) and Bcp3 (*sso0225*) resemble the response of *S. solfataricus* to the infection by SIRV2, which is able in this way to circumvent host defences (Okutan et al., 2013). Expression remodulation of genes involved in informational processing (i.e. transcription and translation) has been also observed for the lytic viruses STIV (Ortman et al., 2008) and SIRV2 (Okutan et al., 2013). This indicates that all these viruses indicate the dependency of SSV2 on the host apparatus and its attempt to take the control over the host informational machinery, despite the difference in their life cycle.

One of the most interesting outcomes from this analysis was the expression variation observed for a group of genes involved in CRISPR-Cas antiviral system, i.e.: i) the type IA interference cassette located downstream the locus C, which includes the genes *sso1439* (*cas3''*), *sso1441* (*cas5*), *sso1442* (*cas7*) and *sso1443* (*csa5*); ii) *sso1424* (small subunit CASCADE complex) and *sso1425* (*csm3*) belonging to the type IIIA interference cassette localized downstream the former one; iii) the *sso1389* (*csx1* putative transcription factor) localized upstream the locus A and iv) *sso1997* (cas7 type IA) lying in the proximity of the locus F (Fig. 2, Tab. 3).

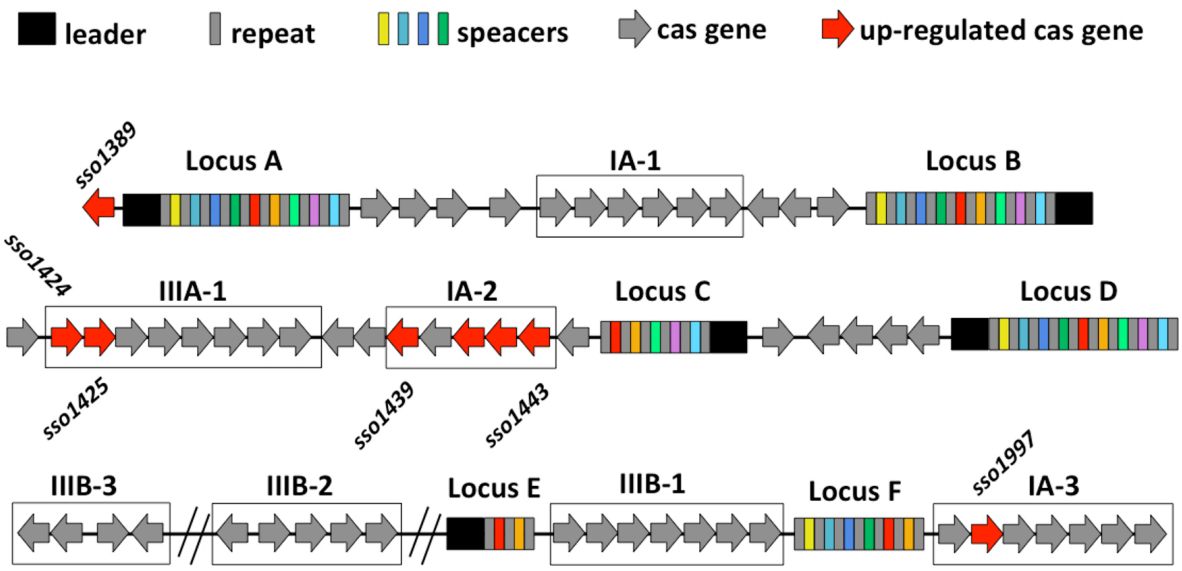


Figure 2. CRISPR/cas loci of *S. solfataricus* P2. The six CRISPR/*cas* loci of the *S. solfataricus* P2 genome are schematized, where red-filled arrows are *cas* genes up-regulated in the SSV2-InF1 stain. Gene cassettes are black-framed and labeled.

Table 3
SSV2-infection affected CRISPR loci and *cas* genes

Gene ID	Function
	CAS GENE PRODUCT
SSO1389	*csx1* (CRISPR-Cas system-associated protein)
SSO1424	*csa5* (CRISPR-Cas system-associated protein)
SSO1425	*csm3* (CRISPR-associated RAMP)
SSO1439	*cas''* (Nuclease subunits of Cas3)
SSO1441	*cas5* (CASCADE complex core)
SSO1442	*cas7* (CRISPR-Associated Protein, Csa2)
SSO1443	csa5 (CRISPR-Associated Protein)
SSO1997	CRISPR-associated auto-regulator, DevR family homolog

CRISPR locus	Probe position
SSOLCTR-A	1708..1767
SSOLCTR-A	3373..3432
SSOLCTR-B	STIV
SSOLCTR-B	278..343
SSOLCTR-B	1493..1557
SSOLCTR-C	pNOB8
SSOLCTR-C	10-71
SSOLCTR-C	1916-1980
SSOLCTR-D	ATV
SSOLCTR-D	105..174
SSOLCTR-D	483..551
SSOLCTR-D	5568..5630
SSOLCTR-F	3689..3753
SSOLCTR-F	4350..4419

In parallel, with the exception of the locus E, all the CRISPR loci were found to be up-regulated during the growth of the SSV2-InF1 strain, with the highest expression induction observed for loci A, C and D (Tab. 3). Noteworthy, proteins encoded by the aforementioned up-regulated type IA interference cassette constitute the CASCADE effector complex of *S. solfataricus* (CRISPR-associated Complex for Antiviral Defence), which is involved in the interference step of the defence system (Zhang and White 2013); thus, hinting to interference processes acting against SSV2. Interestingly, loci A, B, D and F of the *S. solfataricus* P2 strain contain spacers matching fuselloviral genomes with the highest density observed at the leader-proximal region of the locus F (Lillestøl et al., 2009). Although this latter is a non-extending locus, due to the lack a leader sequence, it can still be useful in challenging viral infection since it generates crRNAs. Accordingly, the aCASCADE (archaeal CASCADE) co-purified with crRNAs from all CRISPR loci (Lintner et al., 2011) and probably exert its activity with all the crRNAs of *S. solfataricus*, including the locus F (Manica and Schleper 2013). Moreover, the expression of this locus has been further confirmed in the strain SSV2-InF1 by RT-PCR (data not shown).

The presence of SSV1 quenches the effect of SSV2 infection on host gene expression

In order to evaluate the global variation of the host genes transcription caused by the co-presence of SSV1 and SSV2, slides hybridizations were carried out using cDNA samples as follows:

SSV1/SSV2-InF1 collected at 0.4 OD_{600nm} (ref)
vs
SSV1/SSV2-InF1 collected at 1.2 OD_{600nm} (exp)

As done before, data from the uninfected strain were used to filter those of the double-infected one, so that gene expression variation reported is merely due to the presence of the viruses. Among viral genes, no significant variation was observed with the exception of the SSV2 gene *a305* that showed a down-regulation trend, similarly to that observed for the single-infected strain SSV2-InF1 (Tab. S4). On the other hand, among 29 regulated host genes, 11 were up- and 19 down-regulated during the growth of the double-infected strain (Tab. 2). Therefore, the presence of SSV1 in the same host cells of SSV2 quenched the gene expression variation from 162 (being differentially regulated in the single-infected SSV2-InF1 strain) to only 30 genes. It is worth of note that genes and clusters of the CRISR/Cas system did not undergo to transcription induction in double-infected cells. Instead, the expression pattern of the viral genes is overall identical to that displayed by the single-infected SSV1- and SSV2-InF1 strains. The two fuselloviruses show syntenic genomes, with the only significant difference in the UV-inducible of SSV1, which is lacking in the SSV2 genome. Therefore, it is tempting to speculate that the SSV1 F55 transcription factor, encoded by this region, interferes with the remarkable remodulation of the host gene expression induced by SSV2.

Screening for single clones from the CRISPR/Cas responsive SSV2-InF1 strain

To investigate on whether the activation of the CRISPR/Cas system (see above) enabled the cells to get rid of SSV2, we performed a screening of the strain SSV2-InF1 for the isolation of cured cells. Therefore, clones showing a decreased viral titre, compared to the SSV2-InF1 initial population, were further characterized at DNA level to evaluate SSV2 content through semi-quantitative PCR. Among these, results are shown for clones 1, 1Q and $1Q_1$ that are representative of the first, second and third round of screening, respectively. As shown in figure 3, the PCR products were analysed for each sample on agarose gel at the 20^{th}, 25^{th} and 30^{th} cycle of amplification; the initial population SSV2-InF1 was used as reference. Although the intensity of the *orc1* signal is nearly identical when the same amplification cycle is considered, the amplification band relative to the viral product *vp3* decreases at each round of screening until it becomes hardly visible for the clone $1Q_1$ (Fig. 3). These data confirm that the progressive decrease of the viral titre observed for 1, 1Q and $1Q_1$ clones (data not shown), is mirrored by a drop of the intracellular SSV2 DNA content (Fig. 3).

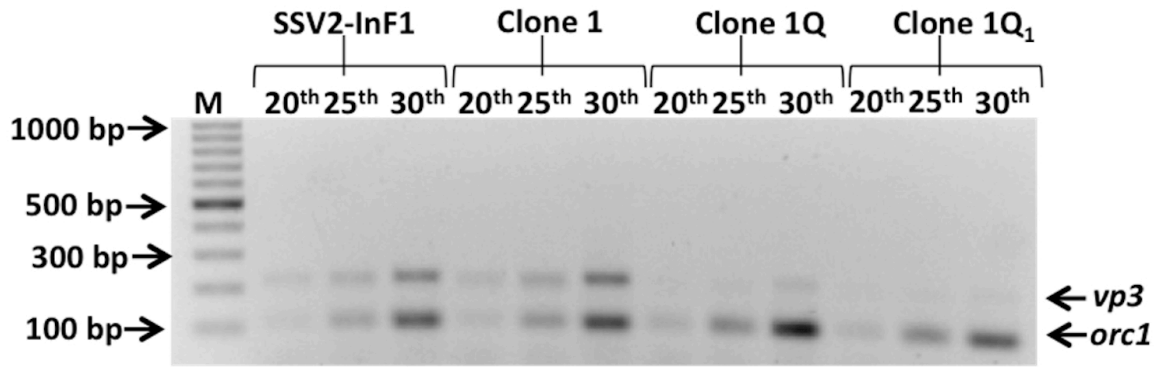


Figure 3. SSV2 DNA detection by semi-quantitative PCR. Black-straight arrows point out to molecular weight markers as well as to the host (*orc1*) and the viral (*vp3*) PCR products. Total DNA samples were prepared from the initial population (SSV2-InF1) as well as from the screened clones (1, 1Q and 1Q). The decreased intensity of the *vp3* amplification band indicates that the SSV2 copy number droped throughout the screenin.

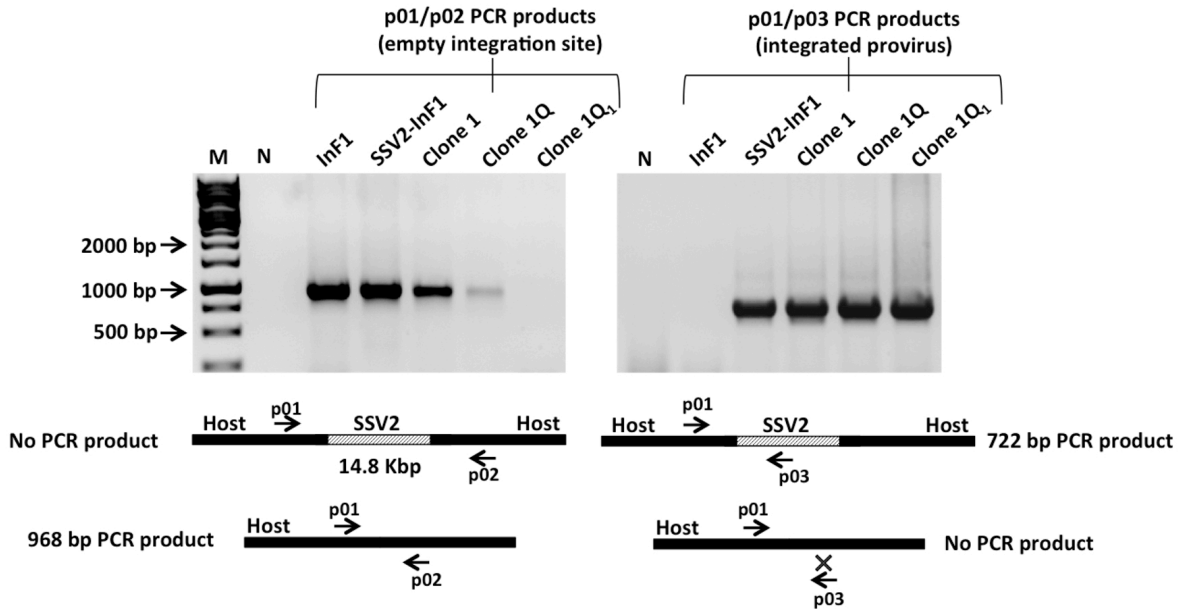


Figure 4. PCR analysis of the integration site occupancy in the screened clones. Black-straight arrows point out to molecular weight markers. PCR products p01/p02 (968 bp, free integration site) and p01/p03 (722 bp, integrated provirus) have been amplified from the SSV2-InF1 as well as from the screened clones (1, 1Q and 1Q), to check the occupancy of the integration site in the host genome. As shown from the right panel, with the exception of the virus-free sample (InF1), all clones carries the provirus. However, provirus-free cells still persist in the cultures (left panel) although their amount decreased throughout the screening, until no empty integration sites are detected after at the third step of screening (1Q1).

In parallel, a PCR analysis of the SSV2 integration site was carried out for the same clones. To do this, the primers SSV2attAp01, SSV2attAp02 and SSV2attAp03 (Contursi et al., 2006) were chosen. The primer couple Ap01-Ap02 amplify a product of about 968bp only if the integration site is empty (no provirus). On the other hand, the couple Ap01-Ap03 lead to the amplification of a 722bp product only if the integration site is occupied by SSV2 (Fig. 4). The presence of the product Ap01-Ap02 for the infected strain SSV2-InF1 as well as for the clone 1 and 1Q indicates that a portion of the cell population do not carry the provirus. Whereas, the absence of such amplicon in the $1Q_1$ clone, suggests that virtually all the cells have the integrated SSV2 in the attachment site. The amplification of the product Ap01-Ap03 further confirms the presence of the provirus (Fig. 4). Since the excision of the provirus strictly depends on the presence of an active integrase/excision enzymes, encoded by the episomal copies (She et al., 2001), the degradation of the latters by the host CRISPR/Cas system led to the entrapment of the integrated viral genome (Fig. 4) as already hypothesized elsewhere (Garret et al., 2011).

SSV2 infection causes deletions at the leader-proximal region of the locus F

A fundamental step of the CRISPR-cas system is the adaptation, which consist in the Cas proteins-mediate incorporation of new spacers from foreign DNAs that occurs immediately downstream the leader sequence of a given CRISPR locus upon viral infection. In order to detect the integration of new spacers, leader-proximal regions of all the six loci in the *S. solfataricus* genome were amplified (Fig. 5). In particular, primer couples were chosen as described elsewhere (Erdman and Garrett 2012), so that integration of new spacers would have produced longer PCR products when compared with a control sample (uninfected InF1).

Our PCR analysis detected no differences in the length of the amplified products (loci A-E) among the initial population SSV2-InF1 infected and the screened clones (Fig. 5), thus confirming that no adaptation occurred upon infection with SSV2 at these loci. Nevertheless, smaller PCR products (between 230 and 450 bp) were found when the leader-proximal region of the locus F was amplified from the $1Q_1$ DNA (Fig. 5). In order to shed light on the identity of these amplicons, PCR products were run on an agarose gel and single bands were cut off, purified and sequenced. Surprisingly, this analysis showed that progressive deletions occurred at the leader-proximal region of the locus F and up to six repeat-spacer units were found to be lacking (Fig. 6).

In this regard, it is noteworthy that Stern and co-workers have point out that the incorporation of self-targeting spacers leads to autoimmunity and selects for the loss of CRISPR functions (Stern et al., 2010). Moreover, it has been demonstrated that when *S. solfataricus* is transformed with a plasmid that carries an essential gene and is also a target of the CRISPR system, surviving transformants show mutations that eliminate the plasmid-targeting spacer (Gudbergsdottir et al., 2011). What would happen if the invading DNA, such as a virus, could not be destroyed because managed to integrate into the host genome, thus becoming ‘self’? In the hypothesis that the host is challenging the SSV2 infection (as shown by microarray data), it is likely that the CRISPR-Cas system targets also the integrated viral genome (i.e., the provirus). This event casts a huge fitness cost on cells harbouring the self-targeting spacer (Dyall-Smith 2011), therefore the selective pressure could drive the inactivation of the CRISPR.Cas system to ensure the survival of SSV2-lysogens. Our results indicate that this is achieved through deletion of spacer units responsible for the recognition of the SSV2 genome (Fig. 5 and 6).

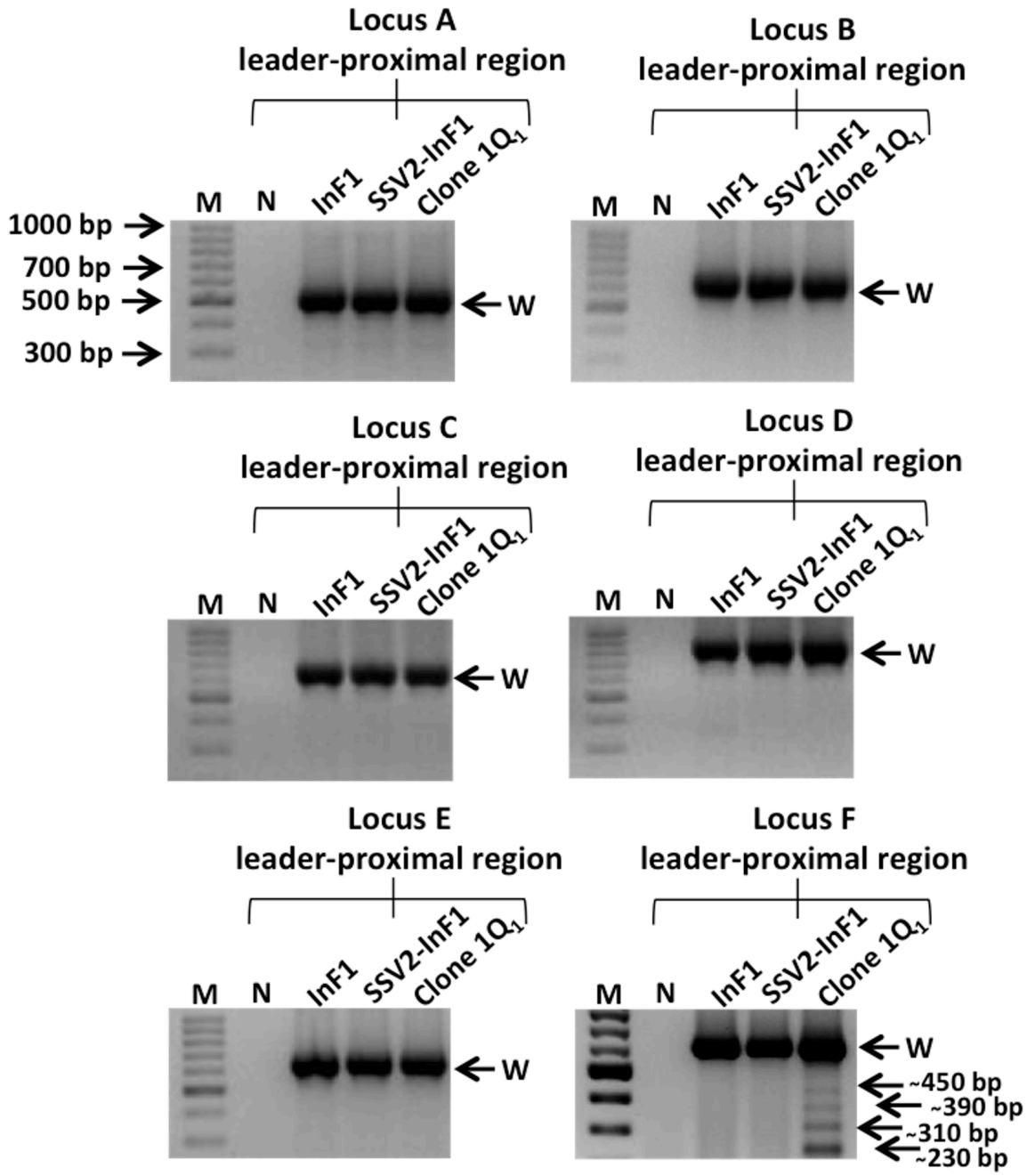


Figure 5. Leader-proximal region amplification of the *S. solfataricus* P2 CRISPR array. PCR amplifications have been performed to amply leader-proximal regions of all six CRISPR loci (A to F) from the virus-free strain (InF1), SSV2-infected initial population (SSV2-InF1) and from the screened clone $1Q_1$. The wild-type PCR product (W) has been amplified for all the loci, thus indication no acquisition of spacers (no adaptation). Intriguingly, shorted PCR products were amplified from the locus F of the clone $1Q_1$. Sequencing of these fragments confirmed loss of repeat/spacer units from this locus (see Fig. 6).

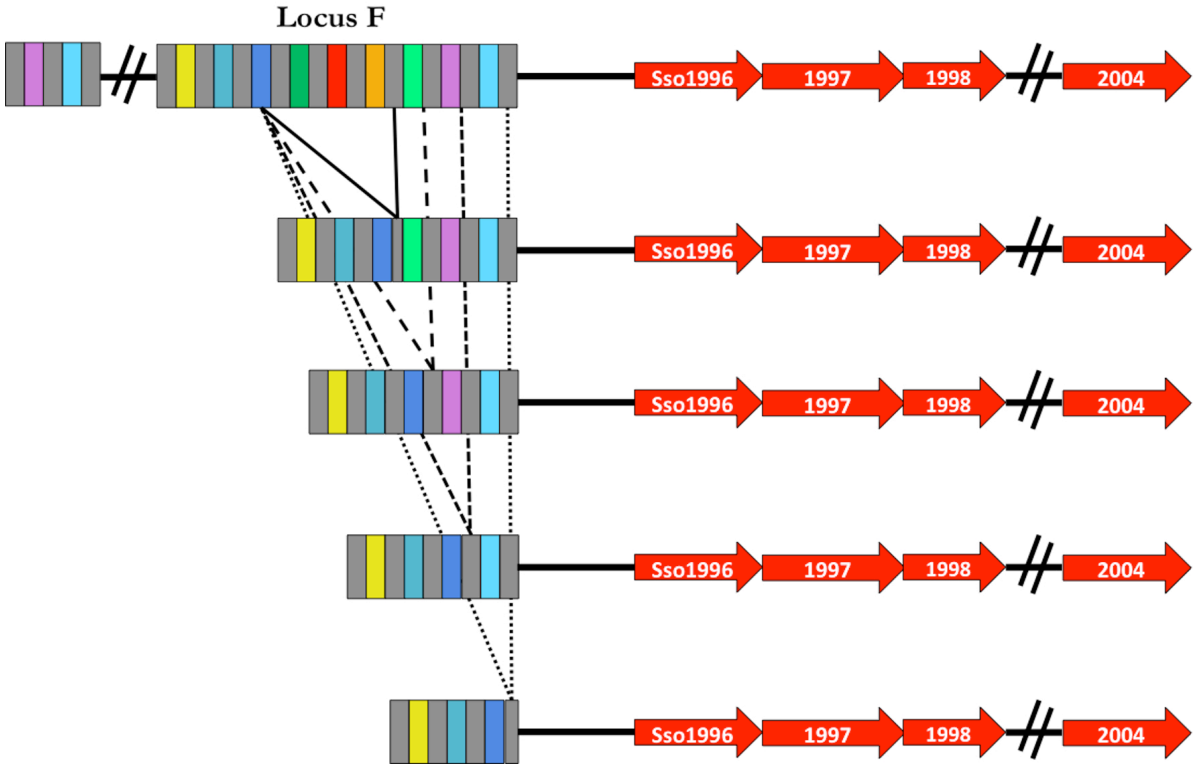


Figure 6. Schematic illustration of repeat/spacer units loss from the locus F. The array F is reported as grey rectangles (repeats) interspersed by colourful rectangles (spacers). A short intergenic region separates the array from the opposite-oriented *cas* genes cassette (IA3, see also Fig. 2). Deletions are schematised to show that from 2 to 6 repeat/spacer units were lost in the clone $1Q_1$.

Conclusions

Analysis of both viral and host transcriptome upon SSV1 and SSV2 infection in *S. solfataricus* has revealed new insights into archaeal host-virus interactions. Previous studies were carried out to detect variation of gene expression during the infection process or upon UV-irradiation. Herein, by exploiting stable infected cells and in the absence of any inductive viral stimulus, we demonstrate not only that the host response towards the two viruses is extremely different during lysogenic growth, but that it is also influenced by the co-presence of both fuselloviruses (Tab. 2). In particular, whereas *S. solfataricus* establishes a harmonious coexistence with SSV1, the reaction against the infection by SSV2 is outstanding and includes the transcriptional activation of CRISPR loci and *cas* genes (Tab. 3). This result is consistent with the isolation of single clones showing low SSV2 viral titre and intracellular copy number (Fig. 3). Our study highlighted an interesting aspect of the host-viral interaction in the frame of the CRISPR-response, i.e. host cells containing an integrated provirus (Fig. 4) are forced to develop a surviving strategy in order to avoid self-attack of the CRISPR-system. In fact, since the provirus is identical in sequence to the episomal copies, it can be targeted by the CRISPR system as well, thus casting a fitness cost on cells that are actively challenging the SSV2 infection. Our data show, for the first time, that *S. solfataricus* cells developed a specific strategy to safeguard host genome integrity, i.e. throughout deletion of self-targeting spacers (Fig. 5 and 6). How deletion of specific spacers at CRISPR loci to avoid self-

attack can occur is matter of further investigation and it is expected to reveal completely novel molecular components and/or mechanism of the complex CRISPR response.

Another relevant aspect revealed by this study is the absence of the CRISPR response in the double-infected SSV1/SSV2 *S. solfataricus* strain (Tab 2 and S4). The two viral genomes are overall identical with the only significant difference being the absence in the SSV2 genome of a region responsible for the UV-induction in SSV1 (Fig. 1). It is tempting to speculate that transcription factors and/or other molecular components encoded by this SSV1 region are responsible for silencing the CRISPR response in the double infected strain. A combination of biochemical and genetic approaches is needed to move toward a better understanding of the mutual influence of the two fuselloviruses in this virus-host system.

Acknowledgements

We thank the *Sulfolobus* Gene Chip Consortium for constructing the microarrays used for our experiments . This study was supported by grants from Danish Independent Research Council (DFF-0602-02196B, DFF-1323-00330 and 11-106683), from the Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica (Progetti di Ricerca di Interesse Nazionale E61J10000020001) and a grant from the Carlsberg Foundation.

References

Ceballos RM, Marceau CD, Marceau JO, Morris S, Clore AJ, Stedman KM. 2012. Differential virus host-ranges of the *Fuselloviridae* of hyperthermophilic *Archaea*: implications for evolution in extreme environments. Front Microbiol. **3**:295.

Comeau AM, Hatfull GF, Krisch HM, Lindell D, Mann NH, Prangishvili D. 2008. Exploring the prokaryotic virosphere. Res Microbiol **159**:306-13.

Contursi P, D'Ambrosio K, Pirone L, Pedone E, Aucelli T, She Q, De Simone G, Bartolucci S. 2011. C68 from the *Sulfolobus islandicus* plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. Biochem J **435:**157-166.

Contursi P, Farina B, Pirone L, Fusco S, Russo L, Bartolucci S, Fattorusso R, Pedone E5. 2014b. Structural and functional studies of Stf76 from the *Sulfolobus islandicus* plasmid-virus pSSVx: a novel peculiar member of the winged helix-turn-helix transcription factor family. Nucleic Acids Res. **42**:5993-6011.

Contursi P, Fusco S, Cannio R, She Q. 2014a. Molecular biology of fuselloviruses and their satellites. Extremophiles **3:**473-489.

Contursi P, Fusco S, Limauro D, Fiorentino G. 2013. Host and viral transcriptional regulators in *Sulfolobus*: an overview. Extremophiles **17**:881-895.

Contursi P, Jensen S, Aucelli T, Rossi M, Bartolucci S, She Q. 2006. Characterization of the *Sulfolobus* host-SSV2 virus interaction. Extremophiles **10**:615-627.

Cooper CR, Daugherty AJ, Tachdjian S, Blum PH, Kelly RM. 2009. Role of vapBC toxin-antitoxin loci in the thermal stress response of *Sulfolobus solfataricus*. Biochem Soc Trans **37**:123-126.

Dyall-Smith M. 2011. Dangerous weapons: a cautionary tale of CRISPR defence. Mol Microbiol **79**:3-6.

Erdmann S, Garrett RA. 2012. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol **85**:1044-1056.

Fröls S, Gordon PM, Panlilioc MA, Schleper C, Sensen CW. 2007 Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology 365:48-59.

Fusco S, Aulitto M, Bartolucci S, Contursi P. 2014. A standardized protocol for the UV induction of *Sulfolobus* spindle-shaped virus 1. Extremophiles. *In press*.

Fusco S, She Q, Bartolucci S, Contursi P. 2013. T_{lys}, a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. J Virol **87:**5926-5936.

Garrett RA, Shah SA, Vestergaard G, Deng L, Gudbergsdottir S, Kenchappa CS, Erdmann S, She Q. 2011. CRISPR-based immune systems of the *Sulfolobales*: complexity and diversity. Biochem Soc Trans. **39**:51-57.

Gudbergsdottir S, Deng L, Chen Z, Jensen JVK, Jensen LR, She Q, Garrett RA. 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol **79**:35-49

Guillière F, Peixeiro N, Kessler A, Raynal B, Desnoues N, Keller J, Delepierre M, Prangishvili D, Sezonov G, Guijarro JI. 2009. Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. J Biol Chem **284:**22222-22237.

Kraft P, Kümmel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM. 2004a. Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol **78:**7438-7442.

Kraft P, Oeckinghaus A, Kümmel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM. 2004b. Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. J Virol **78:**11544-11550.

Krupovic M, Quemin ER, Bamford DH, Forterre P, Prangishvili D. 2013. Unification of the globally distributed spindle-shaped viruses of the *Archaea*. J Virol **88:**2354-2358.

Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, Garrett RA. 2009. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Mol Microbiol **72**:259-272.

Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, Sdano M, Peng N, She Q, Copié V, Young MJ, White MF, Lawrence CM. 2011. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). J Biol Chem **286**:21643-21656.

Manica A, Schleper C. 2013. CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus *Sulfolobus*. RNA Biol **10**:671-678.

Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzer E, Young MJ, Bothner B, Lawrencem CM. 2008. Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology **376:**270-278.

Nadal M, Mirambeau G, Forterre P, Reiter WD, Duguet M. 1986. Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature **321**:256-258.

Okutan E, Deng L, Mirlashari S, Uldahl K, Halim M, Liu C, Garrett RA, She Q, Peng X. 2013. Novel insights into gene regulation of the rudivirus SIRV2 infecting *Sulfolobus* cells. RNA Biology **10**:875-885.

Ortmann AC, Brumfield SK, Walther J, McInnerney K, Brouns SJ, van de Werken HJ, Bothner B, Douglas T, van de Oost J, Young MJ. 2008. Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with Sulfolobus turreted icosahedral virus. J Virol **82**:4874-4883.

Peng X, Garrett RA, She Q. 2012**.** Archaeal viruses-novel, diverse and enigmatic. Sci China Life Sci **55**:422-433.

Pietilä MK, Demina TA, Atanasova NS, Oksanen HM, Bamford DH. 2007. Archaeal viruses and bacteriophages: comparisons and contrasts. Trends Microbiol **22**:334-344.

Pina M, Bize A, Forterre P, Prangishvili D. 2011. The archeoviruses. FEMS Microbiol Rev **35**:1035-1054.

Prangishvili D. 2013. The wonderful world of archaeal viruses. Annu Rev Microbiol **8**:565–585.

Quax TE, Voet M, Sismeiro O, Dillies MA, Jagla B, Coppée JY, Sezonov G, Forterre P, van der Oost J, Lavigne R, Prangishvili D. Massive activation of archaeal defense genes during viral infection. J Virol **87**:8419-8428.

Reiter WD, Palm P, Zillig W. 1988a. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res **16**:1-19.

Reiter WD, Palm P, Zillig W. 1988b. Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. Nucleic Acids Res **16**:2445-2459.

Ren Y, She Q, Huang L. 2013. Transcriptomic analysis of the SSV2 infection of *Sulfolobus solfataricus* with and without the integrative plasmid pSSVi. Virology. **441**:126-134.

Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakić M, Young MJ, Lawrence CM, Copié V. 2012. Structural studies of E73 from a hyperthermophilic archaeal virus identify the ‘‘RH3’’ domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. Biochemistry **51:**2899-2910.

Schleper C, Kubo K, Zillig W. 1992. The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sc USA **89**:7645-7649.

She Q, Peng X, Zillig W, Garrett RA, 2001. Gene capture events in archaeal chromosomes. Nature 409, 478.

Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet. **26**:335-340.
Suttle CA. 2007. Marine viruses–major players in the global ecosystem. Nat Rev Microbiol **5**:801-12.

Wang H, Peng N, Shah SA, Huang L, She Q. 2015. Archaeal extrachromosomal genetic elements. Mol Micro Biol Revs *In press*.

Zhang J, White MF. 2013. Hot and crispy: CRISPR-Cas systems in the hyperthermophile *Sulfolobus solfataricus*. Biochem Soc Trans. **41**:1422-1426.

Supplementary material

Table S1. Uninfected *S. solfataricus* lnF1 strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
NT02SS2357	2.89	0.71	7.43	Intergenic region
SSO0127	1.07	0.17	2.09	2-isopropylmalate synthase
SSO0389	2.24	0.28	4.73	Glycosylated S-layer protein
SSO0488	2.06	1.23	4.18	Phosphate ABC transporter
SSO0489	5.06	0.40	33.46	Phosphate binding periplasmic protein
SSO0490	3.87	0.40	14.64	Phosphate transport system permease
SSO0491	2.51	1.13	5.71	Phosphate transport system permease
SSO0545	2.47	1.20	5.54	Hypothetical protein
SSO0786	2.54	0.83	5.80	Amino acid specific permease
SSO0929	1.56	0.32	2.94	Ribonucleotide reductase (nrd)
SSO1005	1.41	0.59	2.65	Hypothetical protein
SSO10788	2.03	0.12	4.09	Bacterial response and virulence
SSO1092	2.18	1.10	4.54	Hypothetical protein
SSO11071	1.08	0.12	2.12	Pyruvate synthase delta chain
SSO1129	1.74	0.41	3.35	Heterodisulfite reductase, subunit B
SSO11385	1.59	0.47	3.00	non-coding RNA homolog

Table S1.1 Uninfected *S. solfataricus* InF1 strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO1171	1.10	0.14	2.14	Maltose ABC transporter
SSO1183	2.76	1.39	6.76	Inorganic phosphate transporter
SSO11855	1.21	0.34	2.32	STIV-infection differentially regulated
SSO1227	2.03	0.90	4.07	Toluene-4-monooxygenase system protein A
SSO1237	1.14	0.11	2.21	First ORF in transposon ISC1491
SSO1287	1.70	0.64	3.25	Hypothetical protein
SSO1288	2.82	0.79	7.08	Hypothetical protein
SSO1320	1.18	0.18	2.27	Hypothetical protein
SSO1526	1.58	0.86	3.00	Pyruvate dehydrogenase, alfa subunit
SSO1665	1.34	0.54	2.54	Cytosine permease
SSO1835	2.40	1.29	5.26	DNA polymerase beta domain
SSO1858	1.74	0.65	3.34	Hypothetical protein
SSO1860	1.47	0.50	2.77	Bacitracin resistance protein
SSO1865	1.32	0.48	2.49	Universal stress protein family
SSO1894	2.73	0.69	6.64	Hypothetical protein
SSO1910	1.87	0.30	3.66	Hypothetical protein
SSO1931	2.07	1.11	4.20	Hypothetical protein
SSO2029	3.00	0.92	7.99	Hypothetical protein
SSO2044	1.71	0.41	3.27	NAD specific glutamate dehydrogenase
SSO2126	1.69	0.60	3.23	L-lactate permease
SSO2476	1.26	0.01	2.40	Metabolite permease, putative
SSO2505	3.49	0.92	11.24	Sugar transport protein
SSO2523	2.35	0.74	5.09	Long-chain-fatty-acid--CoA ligase
SSO2629	2.03	0.44	4.08	Oxidoreductase (flavoprotein)
SSO2630	1.17	0.38	2.25	SirA family protein
SSO2632	1.44	0.38	2.71	Hypothetical protein
SSO2636	2.27	0.67	4.83	Carbon monoxide dehydrogenase
SSO2756	1.38	0.17	2.60	Pyruvate synthase beta chain
SSO2797	1.18	0.34	2.27	Conserved hypothetical protein
SSO2815	1.26	0.27	2.39	2-oxoacid--ferredoxin oxidoreductase
SSO2863	1.82	1.01	3.53	Acetyl-CoA synthetase
SSO2864	1.29	0.09	2.45	Conserved hypothetical protein
SSO2966	1.36	0.59	2.57	Conserved hypothetical protein
SSO2967	3.04	0.32	8.20	Hypothetical protein
SSO2969	1.53	0.37	2.89	Quinol oxidase-2, subunit I
SSO2970	2.23	0.76	4.69	Quinol oxidase-2, cytochrome b
SSO2971	1.90	0.57	3.72	Quinol oxidase-2, rieske iron-sulfur protein-2
SSO2973	2.61	0.25	6.11	Quinol oxidase-2, subunit I/III, cytochrome aa3
SSO3043	3.35	0.43	10.19	ABC transporter, binding protein
SSO3053	4.67	2.27	25.53	Maltose ABC transporter
SSO3058	2.08	1.66	4.24	Maltose ABC transporter,
SSO3066	2.13	0.57	4.39	Arabinose ABC transporter

Table S1.2 Uninfected *S. solfataricus* InF1 strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO3120	3.28	0.46	9.71	Metabolite transport protein, putative
SSO3180	2.42	0.03	5.33	Phosphate transporter related protein
SSO7127	1.79	0.71	3.46	non-coding RNA homolog
SSO0091	-2.22	0.69	-4.67	LSU ribosomal protein L7AE (rpl7AE)
SSO0164	-1.58	0.29	-2.99	LSU ribosomal protein S8E (rps8E)
SSO0200	-1.29	0.46	-2.44	ArsR family transcriptional regulators homolog
SSO0219	-1.55	0.63	-2.93	SSU ribosomal protein S12AB (rpS12AB)
SSO0278	-1.25	0.38	-2.38	Proteasome subunit
SSO0366	-1.49	0.49	-2.81	Glutamine synthetase
SSO0536	-1.86	0.46	-3.64	S-adenosylmethionine decarboxylase
SSO0742	-1.18	0.26	-2.26	LSU ribosomal protein L15E
SSO0746	-1.17	0.14	-2.26	SSU ribosomal protein S3AE (rps3AE)
SSO0752	-1.24	0.28	-2.37	LSU ribosomal protein L21E (rpl21E)
SSO0962	-1.71	0.39	-3.27	DNA binding protein SSO10b (Alba 1)
SSO1067	-1.34	0.44	-2.54	Second ORF in transposon ISC1359
SSO1100	-1.20	0.33	-2.30	Hypothetical protein
SSO1101	-1.39	0.54	-2.62	Transcriptional regulator
SSO1107	-1.28	0.39	-2.43	Hypothetical protein
SSO1586	-1.65	0.36	-3.13	Conserved hypothetical protein
SSO1593	-2.13	1.10	-4.38	Benzoate transporter
SSO2574	-2.56	0.84	-5.91	Ferredoxin (zfx-1)
SSO2595	-1.07	0.22	-2.10	Conserved hypothetical protein
SSO2613	-1.68	0.04	-3.21	Peroxiredoxin, bacterioferritin comigratory
SSO3155	-1.99	1.04	-3.98	Tryptophan repressor binding protein (wrbA)
SSO5343	-1.02	0.09	-2.02	Conserved hypothetical protein
SSO5410	-1.06	0.05	-2.08	Small nuclear riboprotein protein (snRNP-1)
SSO6454	-1.05	0.12	-2.08	snRNP-2 small nucleare riboprotein
SSO7114	-1.06	0.11	-2.08	SSU ribosomal protein S27E
SSO9724	-1.00	0.11	-2.00	Hypothetical protein

Table S2. *S. solfataricus* SSV1-InF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
small-RNA sR116	1.22	0.31	2.33	small cytoplasmic RNA
SSO0269	1.39	0.54	2.61	GTP-binding protein (hflX)
SSO0276	1.40	0.42	2.64	Small nuclear ribonucleoprotein (snRNP)
SSO0794	1.33	0.10	2.51	Transposase ISC1476
SSO0999	1.45	0.63	2.73	ABC Transporter
SSO10452	2.07	0.67	4.20	Hypothetical protein
SSO1052	1.38	0.39	2.60	Oligosaccharyl transferase STT3
SSO1069	1.85	0.77	3.62	Amino acid transporter
SSO1126	1.34	0.20	2.54	ChainA, Putative Oxidoreductase
SSO1284	1.56	0.36	2.94	Oligo/dipeptide transport, permease protein
SSO1717	1.47	0.66	2.76	Second ORF in transposon ISC1048
SSO1752	1.02	0.02	2.02	Hypothetical protein
SSO2660	1.39	0.55	2.61	Rieske iron-sulfur protein-1 (soxL-1)
SSO2661	1.50	0.29	2.83	P-aminobenzoate N-oxygenase
SSO3085	1.15	0.36	2.22	Hypothetical sulphur transporter
SSO3123	0.93	0.05	1.91	Hypothetical protein
SSO3211	1.02	0.03	2.03	4-aminobutyrate aminotransferase
small-RNA sR102	-1.27	0.53	-2.42	small cytoplasmic RNA
SSO0221	-1.30	0.15	-2.47	LSU ribosomal protein L30E (rpl30E)
SSO0637	-1.74	0.65	-3.34	Hypothetical protein
SSO0961	-1.16	0.23	-2.24	Hypothetical protein
SSO12181	-0.96	0.08	-1.94	Hypothetical protein
SSO2138	-1.53	0.33	-2.89	PadR family transcripitonal regulator
SSO2180	-1.35	0.49	-2.55	Hypothetical protein
SSO2253	-1.12	0.03	-2.18	Rubrerythrin (oxidative stress tolerance)
SSO2381	-0.92	0.05	-1.90	Translation initiation factor 2, beta subunit
SSO2827	-1.09	0.21	-2.13	Predicted transcriptional regulator
SSO5478	-1.05	0.21	-2.07	LSU ribosomal protein L24E (rpl24E)
SSO5663	-1.18	0.25	-2.26	Protein transport protein sec61
SSO5668	-0.98	0.13	-1.97	LSU ribosomal protein LX

Table S3. *S. solfataricus* SSV2-InF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
small-RNA_sR103	1.03	0.10	2.05	RNA
SSO0044	1.29	0.12	2.45	Heme bearing subunit I of the terminal oxidase (doxB)
SSO0045	1.32	0.07	2.50	Terminal oxidase, subunit (doxC)
SSO0227	1.40	0.17	2.63	DNA-directed RNA polymerase, subunit B'' (rpoB1)
SSO0269	1.29	0.20	2.45	GTP-binding protein (hflX)
SSO0276	1.46	0.24	2.75	Small nuclear ribonucleoprotein (snRNP) homolog
SSO0931	1.42	0.12	2.67	Predicted transcriptional regulator
SSO0944	1.04	0.07	2.05	Hypothetical protein
SSO0980	1.30	0.05	2.47	Transposase ISC1217
SSO0999	1.58	0.22	2.99	ABC transporter
SSO1000	1.09	0.21	2.13	Maltose transport inner membrane protein
SSO1004	1.42	0.21	2.67	FAD linked oxidase homologue
SSO10449	1.03	0.07	2.04	Predicted transcriptional regulator
SSO10452	1.24	0.09	2.36	Hypothetical protein
SSO11076	1.44	0.12	2.71	Glycosyltransferase, putative
SSO1125	1.05	0.21	2.08	DsrE/DsrF-like family homologue
SSO1141	1.43	0.36	2.69	Copper binding protein
SSO1152	1.05	0.09	2.07	Peptidase U62 modulator of DNA gyrase homolog
SSO11550	1.27	0.19	2.42	STIV infection differentially regulated gene
SSO11575	1.41	0.19	2.65	Hypothetical protein
SSO1162	1.52	0.13	2.86	Multidrug resistance protein
SSO1268	1.05	0.21	2.07	Glutamine-fructose-6-phosphate transaminase (isomerizing glmS-2)
SSO1281	1.15	0.27	2.21	Last part of transposase in ISC1250
SSO1282	1.93	0.31	3.81	Oligo/dipeptide transport, ATP binding protein . (dppD-2)
SSO1284	2.36	0.28	5.14	Oligo/dipeptide transport, permease protein (dppB-2)
SSO1312	1.75	0.23	3.36	Ring oxydation complex/ phenylacetic acid degradation related protein
SSO1334	1.63	0.29	3.10	Malate synthase, putative (aceB/mas)
SSO1389	1.61	0.34	3.05	csx1(CRISPR/cas system-associated protein)
SSO1439	0.80	0.06	1.74	cas'' (Nuclease subunits of Cas3)
SSO1424	1.45	0.24	2.74	csa5(CRISPR/cas system-associated protein so-called) 'small' subunit
SSO1425	1.21	0.22	2.32	csm3 (CRISPR-associated RAMP)
SSO1441	1.42	0.26	2.67	cas5 (CASCADE complex core)
SSO1442	2.99	0.01	7.93	cas7 (Crispr-Associated Protein, csa2)

Table S3.1 *S. solfataricus* SSV2-InF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO1443	2.56	0.15	5.91	csa5 (CRISPR/cas system-associated protein)
SSO1691	1.16	0.12	2.23	Hypothetical protein
SSO1717	1.42	0.08	2.68	Second ORF in transposon ISC1048
SSO1812	1.85	0.27	3.60	SirA family protein homologue
SSO1814	1.16	0.07	2.23	Transposase ISC1234
SSO1842	1.35	0.49	2.55	Glyceraldehyde-3-phosphate dehydrogenase, NADP dependent (gapN-2)
SSO1870	1.85	0.11	3.60	Rusticyanin homologue
SSO1911	0.99	0.06	1.98	Hemerythrin
SSO1997	2.13	0.47	4.37	CRISPR-associated autoregulator, DevR family homolog
SSO2048	1.97	0.19	3.93	Hypothetical protein
SSO2059	1.25	0.38	2.38	Acetyl-CoA synthetase (acetate-CoA ligase acsA-6)
SSO2067	1.25	0.48	2.38	Indolepyruvate ferredoxin oxidoreductase alpha subunit (iorA)
SSO2131	1.30	0.34	2.46	Transcriptional regulator homolog (Lrp/AsnC family)
SSO2291	3.00	0.00	7.98	Protein kinase, putative
SSO2617	1.04	0.11	2.06	Dipeptide ABC transporter permease protein dppC-3
SSO2656	0.99	0.04	1.99	Quinol oxidase (SoxABC), cytochrome B subunit (soxC)
SSO2657	1.41	0.10	2.66	Quinol oxidase (SoxABC), cytochrome aa3 subunit (soxB)
SSO2660	1.23	0.09	2.34	Rieske iron-sulfur protein-1 (soxL-1)
SSO2661	1.22	0.08	2.33	Hypothetical protein
SSO2681	1.00	0.07	2.01	Hypothetical protein
SSO2690	2.11	0.15	4.30	Sulfolobus mercury resistance protein, MerI (Sulfolobus islandicus)
SSO2757	2.60	0.31	6.06	Pyruvic-ferredoxin oxidoreductase alpha chain (porA-2)
SSO2758	1.22	0.16	2.33	Pyruvic-ferredoxin oxidoreductase gamma chain (porG-2)
SSO2761	1.07	0.15	2.09	Acyl-CoA dehydrogenase (acd-5)
SSO2827	1.11	0.07	2.15	Predicted transcriptional regulator
SSO2831	1.06	0.03	2.08	AAA family ATPase
SSO2846	1.21	0.09	2.32	Extracellular ligand-binding receptor homolog
SSO2881	1.09	0.23	2.14	Fe-S oxidoreductase homologue
SSO2882	1.41	0.26	2.65	Transposase ISC1234
SSO2986	1.12	0.02	2.18	Predicted transcriptional regulator
SSO2991	2.00	0.18	4.01	Hypothetical protein
SSO3045	1.23	0.20	2.34	ABC transporter, ATP binding protein
SSO3054	1.76	0.40	3.38	Terminal oxidase, small hydrophobic subunit (doxE)
SSO3107	1.21	0.13	2.31	Dihydroxy-acid dhydratase ilvD

Table S3.2 *S. solfataricus* SSV2-lnF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO3123	1.38	0.17	2.60	Hypothetical protein
SSO3130	1.15	0.12	2.21	Iron-sulfur protein, conserved putative
SSO3131	1.19	0.12	2.28	Heterodisulfide reductase related protein
SSO3174	1.15	0.13	2.23	Histidine kinase (Sulfolobus solfataricus)
SSO3178	1.32	0.04	2.50	Hypothetical protein
SSO3188	1.64	0.11	3.12	Hypothetical protein
SSO5027	1.08	0.12	2.11	Hypothetical protein
SSO5098	1.42	0.15	2.67	Terminal oxidase, small hydrophobic subunit (doxE)
SSO7239	1.13	0.14	2.18	Conserved hypothetical protein
SSO9088	1.64	0.19	3.11	Intergenic region
SSO9180	1.01	0.02	2.02	7 KD DNA-binding protein
SSO9500	1.37	0.14	2.58	Conserved hypothetical protein
SSOLCTR1a_6470_1708-1767	1.07	0.00	2.11	CRISPR Locus A
SSOLCTR1a_6470_3373-3432	0.03	0.03	1.02	CRISPR Locus A
SSOLCTR1a_6470_6329-6394	0.68	0.02	1.60	CRISPR Locus A
SSOLCTR1b_5964_278-343	0.77	0.05	1.70	CRISPR Locus B
SSOLCTR1b_5964_1493-1557	0.63	0.01	1.55	CRISPR Locus B
SSOLCTR1b_5964	0.52	0.04	1.44	CRISPR Locus B
SSOLCTR1c_1993_10-71	1.41	0.08	2.66	CRISPR Locus C
SSOLCTR1c_1993	1.15	0.18	2.22	CRISPR Locus C
SSOLCTR1c_1993_1916-1980	0.94	0.01	1.91	CRISPR Locus C
SSOLCTR1d_6005_483-551	1.40	0.08	2.65	CRISPR Locus D
SSOLCTR1d_6005	0.93	0.17	1.91	CRISPR Locus D
SSOLCTR1d_6005_105-174	0.85	0.16	1.80	CRISPR Locus D
SSOLCTR1d_6005_5568-5630	1.27	0.02	2.42	CRISPR Locus D
SSOLCTR1e_402	0.03	0.05	1.02	CRISPR Locus E
SSOLCTR1f_5776_3689-3753	0.68	0.09	1.60	CRISPR Locus F
SSOLCTR1f_5776_4021-4090	0.00	0.04	1.00	CRISPR Locus F
SSOLCTR1f_5776_4350-4419	0.62	0.19	1.54	CRISPR Locus F
SSO0067	-1.65	0.15	-3.13	SSU ribosomal protein S2AB (rps2AB)
SSO0074	-1.09	0.13	-2.13	SSU ribosomal protein S31AB

Table S3.3 *S. solfataricus* SSV2-lnF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO0098	-1.19	0.16	-2.28	Methionine aminopeptidase 2
SSO0099	-1.70	0.06	-3.24	Conserved hypothetical protein
SSO0173	-1.24	0.01	-2.36	Aspartyl-tRNA synthetase (aspS)
SSO0192	-1.64	0.04	-3.12	Glutaredoxin related protein
SSO0199	-2.21	0.07	-4.64	S-adenosylmethionine synthetase (maT)
SSO0202	-1.31	0.23	-2.48	D-arabino 3-hexulose 6-phosphate formaldehyde lyase (hpS-2)
SSO0217	-1.19	0.26	-2.28	SSU ribosomal protein S7AB (rpS7AB)
SSO0230	-1.49	0.18	-2.81	Nucleoside diphosphate kinase (NDP kinase) (ndk)
SSO0235	-1.05	0.02	-2.08	Radical SAM protein homolog
SSO0240	-1.26	0.22	-2.39	Adenylosuccinate lyase (adenylosuccinase)(ASL) (purB)
SSO0256	-1.11	0.02	-2.16	Conserved hypothetical protein
SSO0292	-1.32	0.10	-2.50	exosome complex RNA-binding protein Csl4 homolog
SSO0320	-1.13	0.20	-2.19	Hypothetical protein
SSO0352	-1.14	0.05	-2.20	Apoptosis-related Tfar19 related protein
SSO0353	-1.39	0.14	-2.63	SSU ribosomal protein S19E (rps19E)
SSO0356	-1.67	0.00	-3.17	Phosphate regulatory protein, putative
SSO0358	-1.11	0.27	-2.16	Conserved hypothetical protein
SSO0363	-1.06	0.16	-2.09	Prolidase
SSO0397	-1.60	0.15	-3.04	Proliferating cell nuclear antigen putative homolog
SSO0408	-1.48	0.09	-2.79	SSU ribosomal protein S13E
SSO0415	-1.54	0.18	-2.91	DNA-directed RNA polymerase, subunit E (rpoE1)
SSO0420	-1.16	0.04	-2.24	Reverse gyrase (topR-1)
SSO0435	-1.71	0.03	-3.28	SSU ribosomal protein S24E (rps24E)
SSO0436	-1.30	0.28	-2.46	Thiazole biosynthetic enzyme
SSO0439	-1.08	0.20	-2.11	tRNA intron endonuclease, putative
SSO0460	-1.62	0.03	-3.08	MRP protein homolog, conserved ATPase (mrp)
SSO0481	-1.16	0.26	-2.23	HAD-superfamily hydrolase homolog
SSO0501	-1.10	0.06	-2.15	Conserved hypothetical protein
SSO0503	-1.23	0.04	-2.35	Conserved hypothetical protein
SSO0553	-1.14	0.17	-2.20	Conserved hypothetical protein
SSO0554	-1.30	0.37	-2.46	Ribosomal protein L11 methyltransferase, putative
SSO0555	-1.31	0.10	-2.48	Conserved hypothetical protein
SSO0608	-1.42	0.37	-2.67	Conserved hypothetical protein
SSO0615	-1.50	0.23	-2.82	Orotate phosphoribosyltransferase (OPRT) (pyrE)

Table S3.4 *S. solfataricus* SSV2-lnF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO0626	-1.96	0.30	-3.88	Phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR synthetase) (purC)
SSO0708	-1.07	0.15	-2.09	LSU ribosomal protein L14AB (rpl14AB)
SSO0757	-1.16	0.06	-2.23	Spermidine synthase
SSO0771	-1.22	0.21	-2.34	Cell division control 6/orc1 protein homolog (cdc6-2)
SSO0881	-1.47	0.01	-2.78	VPS24 Conserved protein implicated in secretion homolog
SSO0886	-1.20	0.12	-2.30	myo-inositol-1-phosphate synthase
SSO0911	-1.19	0.01	-2.28	Cell division protein
SSO0946	-1.89	0.09	-3.70	Transcription initiation factor IIB (TFIIB) homolog (TFB-2)
SSO0951	-1.37	0.19	-2.58	TATA box binding protein, hypothetical (tfIID)
SSO0981	-1.21	0.07	-2.31	Pyruvate kinase (pyK)
SSO10285	-1.35	0.15	-2.55	Coenzyme PQQ synthesis protein
SSO12199	-1.08	0.06	-2.11	Nucleotide pyrophosphohydrolase [Sulfoloblfataripyrophosphohydrolase [Sulfolobus solfataricus]
SSO1889	-1.66	0.24	-3.16	ATP-dependent RNA helicase
SSO1890	-1.15	0.10	-2.23	major facilitator superfamily MFS_1 homolog
SSO2089	-1.07	0.14	-2.09	TenA family transcriptional regulator homolog
SSO2146	-1.28	0.03	-2.42	major facilitator superfamily MFS_1 homolog
SSO2182	-1.13	0.14	-2.19	Isocitrate dehydrogenase, probable (idh)
SSO2190	-1.19	0.10	-2.29	Hypothetical protein
SSO2222	-1.08	0.23	-2.11	Thioredoxin reductase (trxB-1)
SSO2231	-1.59	0.33	-3.00	Hypothetical protein
SSO2255	-1.15	0.05	-2.22	Peroxiredoxin, bacterioferritin comigratory protein homolog (bcp-3)
SSO2279	-1.69	0.06	-3.23	Hypothetical protein
SSO2292	-2.11	0.31	-4.31	Amino acid transporter related protein
SSO2390	-1.57	0.12	-2.98	Inorganic pyrophosphatase, putative (ppa)
SSO2407	-1.36	0.02	-2.56	2-isopropylmalate synthase, putative (leuA-3)
SSO2423	-1.01	0.05	-2.02	Type I phosphodiesterase/nucleotide pyrophosphatase homolog
SSO2431	-1.23	0.20	-2.35	Carbon monoxide dehydrogenase subunit G homolog
SSO2433	-1.09	0.16	-2.12	Carbon monoxide dehydrogenase, small chain (cutC-1)
SSO2583	-1.22	0.04	-2.33	Sulfolipid biosynthesis protein (sqdB)
SSO2585	-1.18	0.26	-2.27	L-lactate dehydrogenase
SSO2598	-1.22	0.15	-2.33	Transcriptional activator (tenA-2)

Table S3.5 *S. solfataricus* SSV2-InF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO2635	-1.12	0.03	-2.17	Hypothetical protein
SSO2653	-1.48	0.08	-2.79	Conserved hypothetical protein
SSO2706	-1.03	0.11	-2.04	5-methylthioadenosine phosphorylase (mtaP)
SSO2778	-1.08	0.06	-2.11	UspA domain protein homolog
SSO3003	-1.14	0.06	-2.21	Glucose 1-dehydrogenase (dhg-1)
SSO3189	-2.58	0.02	-6.00	Amino acid transporter related protein
SSO3200	-1.11	0.07	-2.15	Conserved hypothetical protein
SSO3219	-1.04	0.18	-2.06	Sugar phosphate nucleotydyl transferase
SSO3224	-1.32	0.07	-2.50	Amino acid transporter related protein
SSO3226	-1.06	0.17	-2.09	Fructose-bisphosphate aldolase homolog
SSO5345	-0.97	0.06	-1.96	Traslation elongation factor EF-1 beta subunit
SSO5478	-1.54	0.18	-2.91	LSU ribosomal protein L24E (rpl24E)
SSO5544	-1.20	0.01	-2.30	Carboxylate-amine ligase
SSO5668	-0.96	0.07	-1.94	LSU ribosomal protein LX
SSO5798	-1.16	0.09	-2.24	DNA-directed RNA polymerase, subunit E (rpoE2)
SSO6264	-1.57	0.26	-2.98	Conserved hypothetical protein
SSO6453	-1.20	0.04	-2.29	LSU ribosomal protein L37E (rpl37E)
SSO6830	-1.23	0.18	-2.35	STIV infection differentially regulated gene
SSV2_A305	-1.54	0.04	-2.91	SSV2 ORF

Table S4 *S S. solfataricus* SSV1/SSV2-InF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Similarly regulated in the SSV1-InF1 infected strain
Similarly regulated in the SSV2-InF1 infected strain
Similarly regulated in both single-infected strains

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
non-coding-RNA_Sso-181	1.07	0.06	2.11	non-coding RNA
SSO0276	1.67	0.06	3.18	Like-Sm ribonucleoprotein core homologue
SSO0445	1.14	0.15	2.20	Agmatinase (agmatine ureohydrolase) (speB-1)
SSO0535	1.04	0.06	2.06	Acyl carrier protein synthase
SSO0544	1.02	0.05	2.03	ExsB family protein homologue
SSO1052	1.16	0.08	2.23	Oligosaccharyl transferase STT3 subunit (B5)

Table S4.1 *S S. solfataricus* SSV1/SSV2-InF1 infected strain: a complete list of the up-regulated and down-regulated genes throughout the growth

Similarly regulated in the SSV1-InF1 infected strain
Similarly regulated in the SSV2-InF1 infected strain
Similarly regulated in both single-infected strains

Gene ID	Average log2 Ratio	S.D. log2 Ratio	VGEF	Function
SSO1284	1.07	0.07	2.10	Oligo/dipeptide transport, permease protein (dppB-2)
SSO1312	1.46	0.27	2.74	Ring oxidation complex/ phenylacetic acid degradation related protein
SSO1859	1.11	0.04	2.16	Heat shock protein (htpX-1)
SSO2048	1.28	0.12	2.44	Hypothetical protein
SSO3051	1.41	0.10	2.66	Alpha-glucosidase (malA)
SSO0099	-1.22	0.15	-2.34	Glutaredoxin related protein, metal-dependent hydrolase
SSO0192	-1.15	0.11	-2.22	Glutaredoxin related protein
SSO0356	-1.37	0.33	-2.58	Phosphate regulatory protein, putative
SSO0397	-1.09	0.10	-2.13	Proliferating cell nuclear antigen putative homologue
SSO0420	-1.07	0.04	-2.10	Reverse gyrase (topR-1)
SSO0437	-1.16	0.17	-2.23	3-octaprenyl-4-hydroxybenzoate decarboxylase
SSO0460	-1.14	0.16	-2.20	MRP protein homologue, conserved ATPase (mrp)
SSO0503	-1.11	0.15	-2.16	Hypothetical protein
SSO0626	-1.34	0.25	-2.53	Phosphoribosylaminoimidazole-succinocarboxamide synthase
SSO0881	-1.19	0.08	-2.28	Hypothetical protein
SSO0951	-1.26	0.11	-2.40	TATA box binding protein, hypothetical (tfIID)
SSO1877	-1.25	0.06	-2.37	Small metal-binding protein
SSO1889	-1.03	0.35	-2.05	ATP-dependent RNA helicase
SSO2292	-1.89	0.30	-3.71	Amino acid transporter, putative
SSO2390	-1.01	0.03	-2.02	Inorganic pyrophosphatase, putative (ppa)
SSO3189	-1.97	0.21	-3.93	Amino acid transporter, putative
SSO5544	-1.23	0.21	-2.35	Carboxylate-ammine ligase
SSO5847	-1.01	0.26	-2.01	Hypothetical protein
SSV2_A305	-1.20	0.26	-2.29	SSV2 ORF

Chapter IV

– Concluding remarks–

4.1 Concluding remarks

Nowadays, many industries take advantage from microbial metabolism to transform raw substrates into value-added products. These established industrial workflows can be ideally divided into: 1) those producing bacterial biomasses as end products (such as starter fermentative cultures, live mucosal vaccines as well as probiotics) and 2) those that use whole-bacteria or microbial-derived biocatalysts for fermented food (wine, beer and cheese) and/or bioactive molecules production (such as enzymes or metabolites). Although microorganisms are valuable bio-factories exploitable for a countless number of applications, they are, at the same time, susceptible to viral infections as the majority of living organisms. It has been demonstrated that viruses can interfere with any industrial workflow by causing loss of microbial biomass and, in turn, leading to a drop of the production yield. As a consequence, companies that rely on bioprocesses invest in research to boost microbial defences. As mentioned before, thanks to their ability of rapidly respond to environmental stimuli, prokaryotes (*Bacteria* and *Archaea*) have evolved a wide 'arsenal' of strategies to defend against viral infections. Potentially, all these mechanisms of resistance could be exploited in order to generate robust microbial strains for biotechnological applications. In this regard, the discovery of the CRISPR-Cas antiviral defence system was considered as a promising solution, because this system is not only able to challenge viral infections but, unlike other defence strategies, it is sequence-specific, adaptive and heritable. This led to foresee the opportunity of 'programming' microbial strains to be selectively immune against viral infections, thus setting up safeguarded bioprocesses through naturally or artificially adding new spacers into CRISPR loci.

Recent comparative genomic analyses have shown a high abundance of genes encoding for defence systems in the genomes of thermophiles. Moreover, it has been observed that these genes cluster into genomic regions referred as defence islands. In parallel, direct analyses of environmental samples have revealed that viruses infecting thermophilic *Archaea*, especially those belonging to the phylum *Crenarchaea*, are extremely different from the bacterial and eukaryal counterparts. These evidences have fostered numerous studies aimed at dissecting host-virus interactions, especially in the frame of the CRISPR-Cas system response. The model archaeon used in this study, i.e. *S. solfataricus* P2, is suitable for this purpose because it shows a complex organization of CRISPR loci and *cas* genes. Additionally, it is a permissive host for SSV1 and SSV2, which belongs to a viral family that groups viruses with morphology, genome organization and genes sequence that have never been observed before.

The first aim of this project was to study the interaction between *S. solfataricus* P2 and the UV-inducible fusillovirus SSV1. This because preliminary whole-transcription analyses revealed that this virus did not induce the activation of the CRISPR-Cas system once infected the host. Therefore, the scope of this analysis was to understand how SSV1 could establish this harmonious relationship with *S. solfataricus*. Interestingly, this led to reveal a new aspect of the SSV1 life cycle regulation. Indeed, a novel viral gene (*f55*) was found to be expressed at high level in SSV1-lysogens, thus indicating an important role of the encode protein (F55) in the maintenance of the virus-host relationship during lysogeny. Given the similarity between the SSV1 genome region harbouring *f55* and that from which the lambda repressor CI is expressed, our interest was to functional characterize the putative transcription repressor F55 in order to unravel the lysogeny/induction switch of SSV1.

The reported experiments confirmed that the RHH DNA-binding protein F55 site-specifically binds *in vitro* to target sequences in the promoters of the UV-inducible T_{ind}, the early-induced T_5, T_6 transcripts as well as to its own promoter. Furthermore, the setting up a suitable protocol for the UV irradiation of the SSV1-lysogens, enabled investigations on the regulative role of F55 *in vivo*. This analysis showed that F55 is indeed the key regulator of the lysogeny/induction transition of SSV1. In particular, it acts by repressing the transcription of SSV1 early-genes in lysogenic cells and by allowing their expression once the infected cell is exposed to UV-light. This makes F55 the first fuselloviral regulator for which a defined role has been proved *in vivo*. Moreover, the analysis of the viral titre and copy number led to highlight substantial differences between the lysogenic state of the phage lambda and that of SSV1. For instance, lambda expresses only the CI repressor during lysogeny and is carried exclusively as a provirus. On the other hand, SSV1 expresses other genes besides F55 that altogether allow the co-presence of both episomal and integrated DNA as well as the production of SSV1 virions throughout the growth of the infected host. As a result, the lysogeny of SSV1 was defined as a carrier stage, in which the virus keeps low titre and copy number. In this latter case, the UV irradiation leads to step into a stage at high copy number that does not cause apparent cell lysis (induced state). These differences are also mirrored by the regulation of the F55 expression, which has been shown not to be as stringent (on/off) as for CI of lambda. In fact, in the case of SSV1, the transition from lysogeny to induction is regulated by oscillation in the F55 intracellular concentration rather than by its presence/absence as reported for CI. Arguably, this leaky regulation evolved as consequence of the non-lytic nature of SSV1, which can produce a viral progeny also in the lysogenic host, without causing cell death. Have the ability of SSV1 to be self-consistent in regulating its copy number been responsible for the defeated activation of the CRISPR-Cas system in *S. solfataricus*? Data collected for the SSV1/SSV2 double-infected strain led to speculate that transcription factors (probably F55) and/or other molecular components encoded by SSV1 might be responsible for direct or indirect silencing the CRISPR-Cas response in the SSV1-infected host. Anyhow, further investigations are underway to define this aspect. For instance, analysis of the leader-proximal region of the *S. solfataricus* CRISPR loci are planned to be carried out in the SSV1-InF1 strain after cycles of UV-irradiations. This would allow understanding if new spacers acquisition is stimulated when SSV1 is present at high copy number for a wider time windows.

On the other hand, as concerns the interaction between *S. solfataricus* and SSV2, reported data clearly shown that this virus triggers a CRISPR-Cas antiviral response once it establishes a stably infection into the host cells. Despite the similarities with SSV1, SSV2 shows a major difference in the regulation of the replication induction. Indeed, in this latter case, it is not an exogenous stimulus that induces viral replication but it is a host-depended signal, which is probably produced when the cells enter into the stationary phase of the growth, at least as concerns the natural host *S. islandicus*. Since in the permissive host *S. solfataricus* it has been shown that this regulation is lost, the stain SSV2-InF1 has provided an exceptional opportunity to study host response toward the infection of a virus that does not self-regulate its replication induction. Almost all six CRISPR loci of *S. solfataricus* were found to be up-regulated in this strain, together with *cas* genes coding for proteins that assemble into a type I interference complex. This suggests that the final target is the SSV2 DNA rather than RNA. However, PCR analysis of the leader-proximal region of all the loci proved that no adaptation occurred as consequence of the SSV2

infection; hence, immunity against SSV2 is probably conferred by previously acquired spacers (Fig. 1, phase I). Even though it was expected that this response would have allowed the host to completely challenge the viral infection, it was found that the virus managed to skip the CRISPR-Cas system. This is supported by the following observations: i) it was impossible to isolate SSV2-cured clones from the SSV2-InF1 strains, although this latter was challenging the infection (Fig. 1, pahse I), ii) clones with a low titre and copy number showed a bias in the attachment-site occupancy; indeed, the integration site always harboured the provirus, indicating that provirus-capture events occurred (Fig. 1, phase II) and iii) these clones underwent to spacer loss at the leader-proximal region of the locus F, which harbours spacers directed against SSV2 (Fig. 1, phase III). These evidences raised interesting questions, such as: i) what would happen if the invading virus cannot be readily destroyed because managed to become 'self', as in the case of SSV2? and ii) which are the consequences of these events on the production of virus-resistant strains for biotechnological applications? It is clear enough that these findings cast doubts on the applicability of such defence system to create safeguarded bioprocesses, at least as concerns integrative viruses. Although strains carrying spacers directed against such viruses could be selected, they would be not stable because of self-attach events that could easily occur against the integrated provirus and, in turn, against the host chromosome (Fig. 1, phase III). A suitable alternative would be the use of CRISPR-Cas systems of type III, which target RNA rather than DNA. However, this would limit their use only for hosts with such subtype of CRISPR-Cas systems. Another possibility would require a deep knowledge of the virus-host interaction that, in turn, would allow the selection of starter strains with mutated attachment sites. Even though the drawback of autoimmunity has been highlighted, these data are interesting from an evolutionary point of view. In fact, as aforementioned, adaptation on one side (host) always stimulates counter-adaptation to the other side (virus). Hence, once again, it is confirmed that: 'the evolutionary arms race is a never-ending story!'

Acknowledgements

I thank Professor Qunxin She, from the University of Copenhagen, for having support my research activity during my master and doctoral projects.

Acknowledgements go also to grants from: i) the Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica (Progetti di Ricerca di Interesse Nazionale E61J10000020001), ii) the Danish Independent Research Council–Technology and Production Sciences (grants: 11-106683, DFF-0602-02196B and DFF-1323-00330), iii) a grant from the Carlsberg Foundation and iv) Università Federico II di Napoli and Compagnia di San Paolo (Programma F.A.R.O, IV tornata). Sulfolobus Gene Chip Consortium is acknowledged as well for constructing the microarrays used in this project.

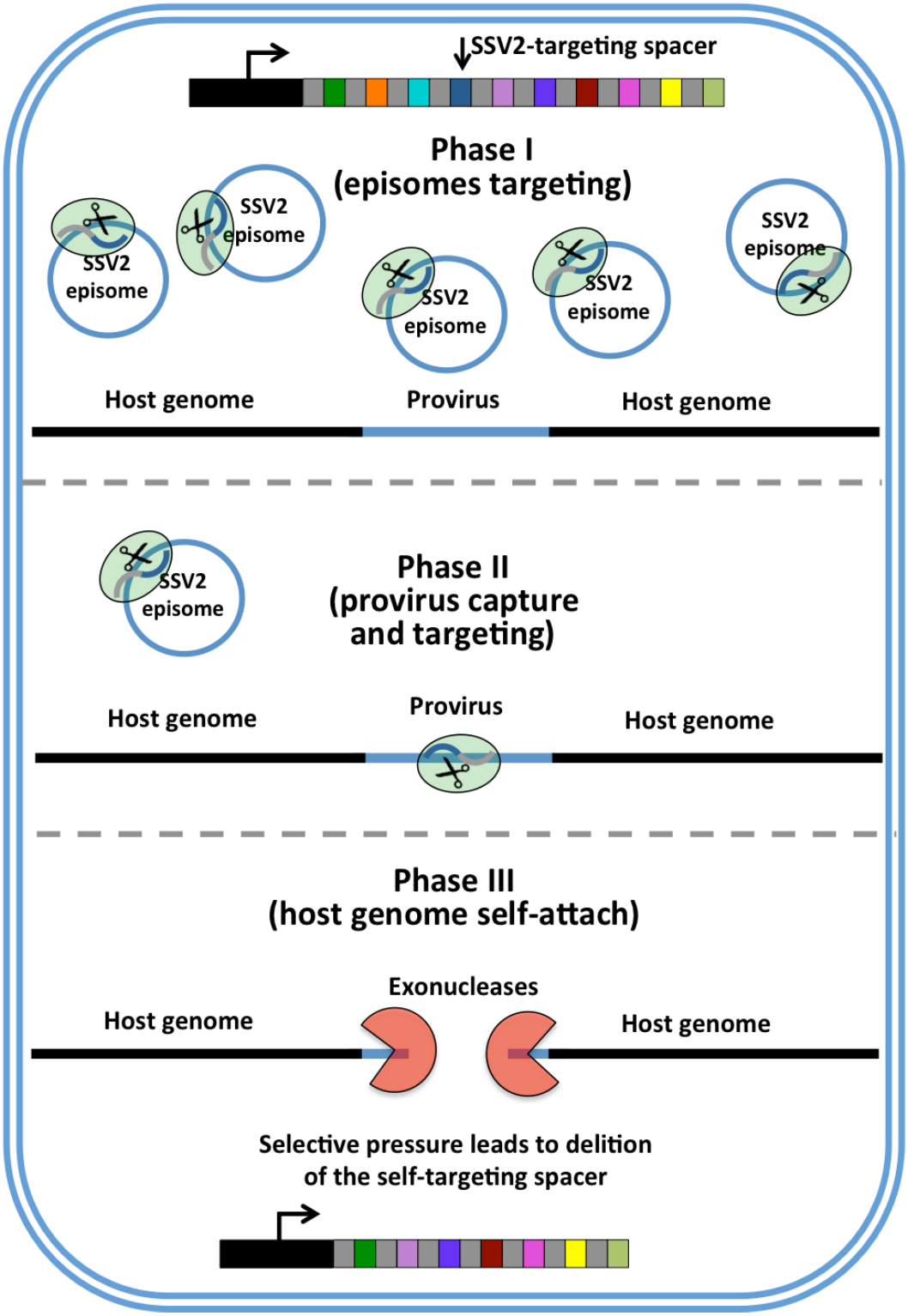


Figure 1. Autoimmunity model. Upon SSV2 infection, targeting-spacers are used to challenge the viral DNA. Since the number of the episomal copies exceeds, it is likely that they are the first to be recognized and targeted (phase I). Once the viral copy number decreases, because of the CRISPR-Cas response, the provirus is trapped into the host chromosome since none or very few functional copies of the integrase genes are present. Moreover, since no differences occur between episomes and provirus, CASCADE complexes can target this latter and cleave it (phase II). This could lead, in turn, to host genome degradation by the action of exonuclease enzymes (Phase III).

Appendices

Appendix I

List of communications

1. **Salvatore Fusco**, Qunxin She, Simonetta Bartolucci and Patrizia Contursi (2012). p55, a newly identified Ribbon-Helix-Helix transcription factor encoded by SSV1: elucidation of archaeal host-virus interactions. Proteine2012, book of abstracts and poster section. Chieti, Italy, 25th-26th September 2012.

2. **Salvatore Fusco**, Qunxin She, Simonetta Bartolucci and Patrizia Contursi (2013). F55: a putative lysogeny regulator of the fusellovirus SSV1. CMC Symposium 2013, book of abstracts. Copenhagen (Denmark), 4th October 2013.

3. Luciano Pirone, Patrizia Contursi, Bianca Farina, **Salvatore Fusco**, Luigi Russo, Simonetta Bartolucci, Roberto Fattorusso and Emilia Pedone (2014). Structural and functional studies of Stf76 from the *Sulfolobus islandicus* plasmid-virus pSSVx: a novel peculiar member of the winged helix-turn-helix transcription factor family. Proteine2014, book of abstracts and poster section. Padova (Italy), 31st March - 1st April 2014.

4. **Salvatore Fusco** (2014). Structure, function and unexplored properties of fuselloviruses-encoded transcription factors. Molecular Biology of *Archaea* 4th conference (**invited speaker**). Institute Pasteur, Paris (France), 19-22 May 2014.

5. Luciano Pirone, Patrizia Contursi, Anna Zanfardino, **Salvatore Fusco**, Mario Varcamonti, Eugenio Notomista, Angela Falanga, Stefania Galdiero and Emilia Pedone (2014). The identification and characterization of a novel CAMP from Stf76, a *Sulfolobus islandicus* plasmid-virus pSSVx transcription factor. 14th Workshop on Bioactive Peptides, book of abstracts and poster section. Naples (Italy), 12nd-14th June 2014.

6. **Salvatore Fusco**, Qunxin She, Simonetta Bartolucci and Patrizia Contursi (2014). Exploring the lysogenic state of *Sulfolobus* spindle-shaped virus 1: the regulative role of the Ribbon-Helix-Helix viral protein F55. 10th International Congress on Extremophiles -Extremophiles2014-, book of abstracts and poster section. Saint Petersburg (Russia), 7th-11st September 2014.

Appendix II

List of publications

1. **Salvatore Fusco**, Qunxin She, Simonetta Bartolucci and Patrizia Contursi (2013). T_{lys}, a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. *Journal of Virology*, **87**:5926-5936.

2. Patrizia Contursi, **Salvatore Fusco**, Danila Limauro and Gabriella Fiorentino (2013). Host and viral transcriptional regulators in *Sulfolobus*: an overview. *Extremophiles*, **17**:881-895.

3. Patrizia Contursi, **Salvatore Fusco**, Raffaele Cannio and Qunxin She (2014). Molecular biology of fuselloviruses and their satellites. *Extremophiles*, **18**:473-489.

4. Patrizia Contursi, Biancamaria Farina, Luciano Pirone, **Salvatore Fusco**, Luigi Russo, Simonetta Bartolucci, Roberto Fattorusso and Emilia Pedone (2014). Structural and functional studies of Stf76 from the *Sulfolobus islandicus* plasmid-virus pSSVx: a novel peculiar member of the winged helix-turn-helix transcription factor family. *Nucleic Acid Research*, **42**:5993-6011.

5. **Salvatore Fusco**, Martina Aulitto, Simonetta Bartolucci and Patrizia Contursi (2015). A standardized protocol for the UV induction of *Sulfolobus* spindle-shaped virus 1. *Extremophiles*, **19**:539-46.

6. **Salvatore Fusco**, Qunxin She, Gabriella Fiorentino, Simonetta Bartolucci and Patrizia Contursi (2015). Unravelling the role of the F55 regulator in the transition from lysogeny to UV induction of *Sulfolobus* spindle-shaped virus 1. *Journal of Virology*, (Under review).

7. **Salvatore Fusco**, Rossana Liguori, Danila Limauro, Simonetta Bartolucci, Qunxin She and Patrizia Contursi (2015). Transcriptome analysis of *Sulfolobus solfataricus* infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system. *Biochemie*,(Under review).

8. Eugenio Notomista, **Salvatore Fusco**, Stefania Galdiero, Emilia Pedone, Luciano Pirone, Mario Varcamonti, Anna Zanfardino and Patrizia Contursi (2015). Identification and characterization of a cryptic antimicrobial peptide in the structure of the transcription factor *Stf76* encoded by the virus satellite pSSVx. (In preparation).

Appendix III

Experiences in foreign laboratories and awards

DEPARTMENT OF BIOLOGY

UNIVERSITY OF COPENHAGEN

Certificate of research stay for Salvatore Fusco at DAC

To Whom It May Concern:

This is to certify that Salvatore Fusco, a PhD student at Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, Naples, Italy, has conducted a research stay at the Danish Archaea Centre, University of Copenhagen for the period of the 2nd of October 2013 to the 22nd of December 2013.

Yours sincerely

Qunxin She

Centre leader,
Associate Professor

2013.12.13

Qunxin She
Centre leader
Danish Archaea Centre

Functional genomics section

BIOCENTER COPENHAGEN

OLE MAALOES VEJ 5

2200 COPENHAGEN N

Phone +45 35 32 20 13

Fax +45 35 32 21 28

qunxin@bio.ku.dk

FEMS Young Scientist Meeting Grant (YSMG)

GRANT CERTIFICATE

FEMS Central Office
Keverling Buismanweg 4
2628 CL Delft
The Netherlands
T: +31-15-269 3920
F: +31-15-269 3921
E: fems@fems-microbiology.org
I: www.fems-microbiology.org

FEMS is honoured to certify that it has provided a grant to

Salvatore Fusco

from

Naples, Italy

in order to attend the FEMS Meeting:

10th International Congress on Extremophiles, Extremophiles 2014

in

Saint Petersburg, Russia

in the period from

07 September 2014 to 11 September 2014

Laura Dib
(FEMS Central Office)

Prof. Dr Jaroslav Spizek
(FEMS Grants Secretary)

Appendix IV

Other publications

Published online 25 March 2014

Nucleic Acids Research, 2014, Vol. 42, No. 9 ***5993–6011***
doi: 10.1093/nar/gku215

Structural and functional studies of Stf76 from the *Sulfolobus islandicus* plasmid–virus pSSVx: a novel peculiar member of the winged helix–turn–helix transcription factor family

Patrizia Contursi[1,†], Biancamaria Farina[2,†], Luciano Pirone[3,†], Salvatore Fusco[1], Luigi Russo[4], Simonetta Bartolucci[1], Roberto Fattorusso[4] and Emilia Pedone[2,5,*]

[1]Dipartimento di Biologia, Università degli Studi di Napoli "Federico II", Napoli 80126, Italy, [2]Interuniversity Centre for Research on Bioactive Peptides (CIRPEB), University of Naples Federico II, Via Mezzocannone 16, 80134 Naples, Italy, [3]Istituto di Cristallografia, C.N.R., Bari 70126, Italy, [4]Dipartimento di Scienze e Tecnologie Ambientali, Biologiche e Farmaceutiche, Seconda Università di Napoli, Caserta 81100, Italy and [5]Istituto di Biostrutture e Bioimmagini, C.N.R., Napoli 80134, Italy

Received December 12, 2013; Revised February 28, 2014; Accepted February 28, 2014

ABSTRACT

The hybrid plasmid–virus pSSVx from *Sulfolobus islandicus* presents an open reading frame encoding a 76 amino acid protein, namely Stf76, that does not show significant sequence homology with any protein with known 3D structure. The recombinant protein recognizes specifically two DNA-binding sites located in its own promoter, thus suggesting an autoregulated role of its expression. Circular dichroism, spectrofluorimetric, light scattering and isothermal titration calorimetry experiments indicated a 2:1 molar ratio (protein:DNA) upon binding to the DNA target containing a single site. Furthermore, the solution structure of Stf76, determined by nuclear magnetic resonance (NMR) using chemical shift Rosetta software, has shown that the protein assumes a winged helix–turn–helix fold. NMR chemical shift perturbation analysis has been performed for the identification of the residues responsible for DNA interaction. In addition, a model of the Stf76–DNA complex has been built using as template a structurally related homolog.

INTRODUCTION

Studies on crenarchaeal viruses have shown that they possess unique morphological features and quasi-orphan genome sequences that distinguish them from bacteriophages and eukaryal viruses (1–3). Indeed, the vast majority of genes from archaeal viruses do not have detectable homologues in the databases other than in other hyperthermophilic viruses (4). Seven families of double-stranded DNA viruses have been identified, among which the *Fuselloviridae* and *Rudiviridae* are the most well-studied specimens and therefore represent model systems for detailed studies of archaeal virus biology (5). These are indeed easily maintained under laboratory conditions and can be obtained in sufficient yields, which is not the case for many other archaeal viruses (6–10). A relatively high proportion of archaeal viral sequences is predicted to carry folds associated to transcription factor (TF) (11). The abundance of putative TFs probably reflects the importance of transcription regulation in the viral life cycle. As in the case of their hosts, the majority of the predicted TFs are bacterial-like and display ribbon–helix–helix, helix–turn–helix (HTH) or looped-hinge helix folds. In addition to bacterial-like TFs, archaeal viruses typically bear one or two sequences with Zinc Finger motifs (4). To date, structural and functional information on archaeal transcription regulators are scarce and only for few of them the 3D structure has been determined (6,8,12–14).

Two distinct genetic elements, SSV2 and pSSVx, belong to *Fuselloviridae* and coexist in the same *Sulfolobus islandicus* REY15/4 host, thus representing one of the few known two-virus systems in Archaea (15). pSSVx is a satellite virus that generates virus particles with the help of SSV2-associated packaging mechanisms. The transcriptional pattern of pSSVx undergoes a temporal variation of gene expression during its own life cycle, thus providing a good model for studying regulation of gene expression in Archaea

*To whom correspondence should be addressed. Tel: +390812534521; Fax: +390812534574; Email: empedone@unina.it
†These authors contributed equally for this work.
The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

(16,17). This genetic element encodes four TFs possibly implicated in the regulation of gene expression, i.e. ORF-c68, ORF51, ORF91 and ORF76. Among these, ORF76, here named Stf76 (*Sulfolobus* TF 76 amino acid protein), has homologs in almost all conjugative and cryptic plasmids from *Sulfolobus* (18), thus suggesting a relevant role for this protein in replication and/or maintenance of the plasmid.

Previous studies showed that transcriptional levels of the *Stf76* gene were constant and fairly high, similar to its homolog ORF80 from the pRN1 plasmid, whose DNA-binding capability has been described (19,20). Nevertheless, unlike ORF80 mRNA, which is synthesized at constant levels up to the stationary/death phase of the host growth (21), the expression of Stf76 transcript (T_{stf76}) was completely inhibited during the process of pSSVx replication induction (16).

In this study, we have performed a detailed structural and functional characterization of Stf76. The corresponding gene has been cloned, expressed in *Escherichia coli* and the recombinant protein purified to homogeneity. To elucidate its interaction with the identified DNA operator sequence, analyses regarding its DNA-binding capabilities by means of electrophoretic mobility shift assay (EMSA), circular dichroism (CD), spectrofluorimetric and isothermal titration calorimetry (ITC) experiments have been performed. Moreover, a structural study has been undertaken by nuclear magnetic resonance (NMR) spectroscopy leading to (i) the solution structure of Stf76 based on CS-Rosetta approach (22,23), (ii) the characterization of the Stf76–DNA interaction by chemical shift perturbation (CSP) analysis, (iii) a structural model describing the interaction of a single Stf76 monomer with its DNA operator. Altogether these results contribute to elucidate the regulatory mechanism underpinning the role of this protein.

MATERIALS AND METHODS

Cloning, expression and purification of Stf76

The *Stf76* gene was polymerase chain reaction (PCR)-amplified from the plasmid pSSVx (cloned in pUC18) by using the primers Stf76fw 5'-CCCTATTTAACATATGG AAAAGGCGAAAC-3' and Stf76rv 5'-CATTACCCCGC TCGAGGTCGGCTAATTCATCTC-3'. The PCR product was digested with *Nde*I and *Xho*I and ligated to the pET-30a(+) digested with same restriction enzymes. Overexpression of Stf76 in the *E. coli* BL21-CodonPlus®(DE3)-RIL cells was induced at OD_{600nm}=0.8 by the addition of 0.5 mM IPTG (isopropyl β-D thiogalactopyranoside) for 16 h. The cells from 1 l of culture were resuspended in 20 ml of lysis buffer (50 mM sodium phosphate buffer, 300 mM NaCl pH 7.0) containing complete protease inhibitor cocktail tablets (Roche). The crude extract was subjected to heat treatment at 70°C for 20 min and then centrifuged for 15 min. The extract was applied to a HisTrap HP (GE Healthcare) equilibrated with buffer A (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole pH 7.0). The column was washed with buffer A plus 30 mM imidazole, proteins were eluted with the same buffer A supplemented with 250 mM imidazole. Protein-containing fractions were pooled and loaded on to a Superdex 75 16/60 column (GE Healthcare) and the elution was carried out in 20 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. Finally, Stf76 was dialyzed overnight against 10 mM sodium phosphate buffer, 50 mM NaCl pH 7.0.

The concentration of the purified protein was measured using the theoretical extinction coefficient $\epsilon = 10\,095\ M^{-1}\ cm^{-1}$ at a Nanodrop 2000 Spectrophotometer (Thermo-Scientific). Alternatively, the Pierce BCA Protein Assay or Coomassie (Bradford) Protein Assay were utilized.

EMSA assay

Binding of Stf76 to target DNA sequences was assayed by EMSA experiments. Two distinct DNA regions (herein named site A and site B), located in the promoter of the *Stf76* gene and previously indicated as putative binding sites of Stf76 (19), were analyzed. The 137 bp DNA fragment encompassing both sites A and B (probe A+B) was amplified by PCR using the pSSVrt (24) as template and the following primers 5′-Stf76 siteA/B (5′-GTTAGCCCACGCG TGAAGGGAAA-3′) and 3′-Stf76 pr siteA/B (5′-AGTTT CGCCTTTTCCATACGTTAAATAGGG-3′). In the displacement experiments, the binding reactions were performed with 3.2 μM Stf76 with the concurrent addition to the EMSA mixtures of increasing amounts of specific unlabeled probe (1:10, 1:100, 1:500, 1:1000 ratio of labeled/unlabeled specific DNA) or Salmon Sperm DNA as aspecific competitor (1:1000, 1:2000 ratio of labeled specific/aspecific DNA).

In order to analyze the binding of Stf76 to each of the two interacting regions, a sequence of 52 bp including only site A (probe A) was PCR-amplified using the following primers 5′-Stf76 siteA/B and Revplr1A(5'-TTCGCCT TCTGAAAATTTGTCTTCATAACAC-3') and used in EMSA assay in comparison with a probe encompassing site B (probe B, 48 bp) obtained with the following primers couple 5'-siteB (5'-GTGTTATGAAGACAAATTTTCAG AAGGCGAA-3')/3′Stf76 siteA/B. A shorter sequence, retaining specific binding of Stf76 to site A, was obtained by annealing a 35-nt-long ^{32}P-labeled oligodeoxynucleotide (probe A') encompassing site A (5′-GGAAACAGTATTA ATAAAGTGTTAATCCTATTACCC-3′) with a complementary oligodeoxynucleotide in a buffer containing 10 mM Tris-HCl pH 7.5, 50 mM NaCl and 1 mM EDTA. Finally, a mutated version of the oligodeoxynucleotide probe A' (probe A'* 5'-GGAAACAGTATTAATAAAG TGCCGTTCCTATTACC-3') was employed to analyze the binding capability of Stf76 to site A.

All the EMSA assays were conducted by a thermal preincubation of the purified Stf76 protein for 15 min at 50°C in assay buffer [20 mM Tris-acetate (pH 8.0), 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT (dithiothreitol) and 5% (v/v) glycerol] in the presence of 1 μg of Salmon Sperm DNA as aspecific competitor and by adding the labeled probes at a concentration of 5–10 nM. The binding reactions were performed with increasing amounts of Stf76 (3.2, 6.4, 9.6, 12.8, 16, 19.2, 22.4, 25.8, 32.4, 40 and 51.6 μM) for 30 min at 37°C and analyzed on 6% or 10% polyacrilamide gels (depending on the size of the probe) in 0.5× TBE.

Gels were transferred onto filter paper, dried and revealed both by Molecular Dynamics Bio-Rad PhosphorImager and/or autoradiography.

CD

Far-UV CD spectra (260–190 nm) were recorded by using a Jasco J-715 spectropolarimeter, equipped with a PTC-423S/15 peltier temperature controller. Spectra were acquired at 20°C according to the following parameters: band width of 1 nm, response of 8 s, data pitch of 0.2 nm and scanning speed of 10 nm/min. CD measurements were carried out using a 0.1-cm-pathlength cell and a protein concentration of 10 μM in a 10 mM sodium phosphate, 50 mM NaCl pH 7.0 buffer. For the titration with the DNA, concentration of the double-stranded oligonucleotides (35 bp-site A) was varied from 0 to 10 μM. The baselines were corrected by subtracting buffer and DNA spectra.

Fluorescence

Stf76 at final concentration of 5 μM in 10 mM sodium phosphate, 50 mM NaCl pH 7.0 buffer was excited at 295 nm and emission spectra among 290 and 450 nm were measured by a spectrofluorimeter (Varian). The spectra were registered also in the presence of increasing concentration of the double-stranded oligonucleotides (35 bp-site A) from 0 to 10 μM.

Thermal denaturation of DNA

Double-stranded DNA (35 bp-site A) at a concentration of 5 μM was incubated in 10 mM sodium phosphate, 50 mM NaCl pH 7.0 with and without 10 μM Stf76 in a 0.5 ml cuvette. The absorbance at 260 nm was monitored while the cuvette was heated with a peltier element from 20 to 80°C, with an increment of temperature of 1°C/min.

Cross-linking

Stf76 (10 μM) was incubated with and without 10 μM double-stranded oligonucleotides (35 bp-site A) in the presence of 100 μM of Bis(sulfosuccinimidyl) suberate (BS3) in 10 mM sodium phosphate, 50 mM NaCl pH 7.0 buffer, for 1 h at room temperature in a reaction volume of 20 μl. The samples were separated on 18% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

Dynamic light scattering and static light scattering

Size measurements of 50 μM Stf76 were performed on a Nano Zetasizer, spectrometer (Malvern, UK) in 10 mM sodium phosphate, 50 mM NaCl pH 7.0 buffer. The wavelength of the laser was 632.8 nm, and the scattering angles 90° and 175°. For each sample, the mean value of particles' diameters was calculated from three replicate determinations. The molecular diameter is calculated from the autocorrelation function of the intensity of light scattered from the particles assuming a spherical form of particles. For molecular weight measurements, a MiniDAWN Treos spectrometer (Wyatt Instrument Technology Corp.) equipped with a laser operating at 658 nm was used connected online to a size-exclusion chromatography. Seventy micromolar Stf76 in 10 mM sodium phosphate, 100 mM NaCl pH 7.0 and 35 μM of double-stranded oligonucleotides (35 bp-site A) was loaded alone and in a molecular ratio of 1:1 and 2:1 on a WTC-015S5 (150 Å) column (Wyatt technology) and analyzed by size-exclusion chromatography connected to a triple-angle light scattering detector equipped with a QELS (Quasi-Elastic Light Scattering) module. A constant flow rate of 0.5 ml/min was applied. Elution profiles were detected by a Shodex interferometric refractometer and a mini Dawn TREOS light scattering system. Data were analyzed by using Astra 5.3.4.14 software (Wyatt Technology).

ITC

ITC studies were performed at 22°C with an iTC200 calorimeter (MicroCal/GE Healthcare, Milano, Italy). Samples were extensively dialyzed in the same buffer (20 mM sodium phosphate, 50 mM NaCl pH 7.0) prior ITC measurements. A solution of Stf76 at a concentration of 100 μM was titrated into a 10 μM double-stranded oligonucleotide 35 bp-site A solution. To exclude the presence of artifacts due to ligand dilution ITC runs were performed by titrating of Stf76 into the buffer. Fitting of data to sequential binding sites model was carried out with the Origin software as supplied by GE Healthcare.

Protein alkylation and LC-ESI-MS analyses

Stf76 (50 μg) was denatured in 50 μl of a solution containing 20 mM sodium phosphate, 6 M guanidinium chloride pH 7.0 for 1 h at 25°C, in the absence of reducing agents. The protein was then subjected to alkylation by incubation of the sample in 0.12 M 4-vinyl-pyridine (4-VP) (Sigma-Aldrich) at 25°C for 60 min, after which time the reaction was terminated by quenching at 4°C. As a control, after denaturation, the protein was reduced before 4-VP alkylation. Then 0.2 μg of alkylated and control sample were run on a Liquid Chromatography-Electrospray-Mass Spectrometry (LC-ESI-MS) instrument. The molecular mass of the protein was estimated using electrospray mass spectra recorded on a Bio-Q triple quadrupole instrument (Micromass, Thermofinnigan, San Jose, CA, USA) as described previously (25).

NMR spectroscopy

All NMR experiments were carried out at 25°C using an Inova 600 MHz spectrometer (Varian Inc., Palo Alto, CA, USA), equipped with a cryogenic probe optimized for ^{1}H detection.

$^{15}N/^{13}C$- and ^{15}N-labeled samples were prepared through a procedure similar to that described above, but cells were grown in M9 medium with the addition of ^{13}C-labeled glucose and ^{15}N-labeled ammonium chloride (26).

Samples used for NMR structural studies contained ~200 μM Stf76 in 20 mM sodium phosphate pH 5.5, 50 mM NaCl, 0.02% sodium azide and 10% $^{2}H_2O$.

Structure determination

Sequential backbone signal assignment was first carried out by combined analyses of the 3D triple-resonance spectra HNCA (27–29) and HN(CO)CA (29,30), and by evaluation of sequential nuclear Overhauser effects (NOEs) in a ^{15}N-edited NOESY spectrum (31–33), acquired with a mixing time of 100 ms. Complete H^N, N, Cα and Hα resonance assignments of all ^{15}N-HSQC-detected residues were subsequently achieved based on ^{15}N-edited TOCSY spectrum (31), acquired with a mixing time of 50 ms. Four glutamine and one asparagine side chain amides were detected and unambiguously assigned in the ^{15}N-edited NOESY spectrum. The indolic HN group of the only Trp residue was also unambiguously assigned. All NMR data were processed with the software VNMRJ 1.1.D (Varian Inc.). 2D and 3D spectra were analyzed using tools available in CARA software (34).

Secondary structural elements of Stf76 were initially identified by chemical shift index (CSI) analysis (35), which was generated using Cα and Hα chemical shifts. Dihedral angle restraints were calculated from H^N, Cα, Hα and N chemical shifts with the software TALOS+ (36).

The structure calculation was performed with the program CS-Rosetta (22,23) using as structural restraints the torsion angles ϕ/ψ derived from TALOS+ database and the H_N, Cα, Hα and N chemical shifts of those residues indicated by TALOS+ to be rigid in pico-second timescale with an order parameter $S^2 > 0.7$. A set of 200 fragment candidates matching these chemical shifts was used to calculate 3000 structures in Rosetta. The energy of these Rosetta structures was then rescored against the observed chemical shifts and the 20 conformers with the lowest rescored energy were selected for the ensemble. The structures were visualized and evaluated by using the programs MOLMOL (37), PROCHECK-NMR (38) and MOLPROBITY (39).

The most representative model of the Stf76 structure was compared to structures in the Protein Data Bank by using the DaliLite server (40).

^{15}N relaxation parameters of Stf76 were evaluated by recording inversion recovery ^{1}H-^{15}N HSQC for R_1 measurements and spin echo ^{1}H-^{15}N HSQC for R_2 measurements . Acquisition parameters and relaxation data processing and analysis were essentially as reported (41,42). An estimate of the isotropic correlation time, τc, of Stf76 was obtained from the ratios of ^{15}N R_2 and R_1 values of residues that passed the coarse and fine-filtering steps (43), using the equation reported (44). The Stokes–Einstein–Debye equation was then used to calculate the hydrodynamic radius from the τc. The hydrodynamic properties were also evaluated using the NMR software hydropro (45).

NMR analysis of Stf76–DNA interaction

CSP studies of Stf76 with the 35-bp-site A sequence were carried out using ^{15}N-labeled Stf76 dissolved in 500 μL at 100 μM concentration in 20 mM sodium phosphate pH 5.5, 50 mM NaCl, 0.02% sodium azide and 10% ^2H_2O. The DNA fragment, prepared as reported above, was first dissolved at 500 μM concentration in 20 mM sodium phosphate pH 5.5 buffer, 50 mM NaCl and then carefully mixed with Stf76 at a final concentration of 10 and 50 μM, as reported in Omicinsky *et al.* (46). 2D [^{15}N, ^{1}H] Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired in the absence and presence of DNA. Starting from the amide resonances for Stf76 free, average combined chemical shift changes for Stf76 bound were determined using the following equation: $\Delta\delta HN_{av} = [((\Delta\delta_H)^2+(\Delta\delta_N/5)^2)/2]^{1/2}$, where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift variations of the amide proton and nitrogen resonances (47–50), respectively. Moreover, intensity reduction of the amide cross-peaks over one standard deviation was also taking in account to define the DNA-binding site.

RESULTS

Cloning, expression and purification of Stf76

The gene *Stf76* was cloned into the expression vector pET30a(+) and the recombinant protein was expressed in BL21(DE3) codon-plus ArglleLeu (RIL) as a soluble C-terminal His-tag fused protein. Stf76 was purified to homogeneity in a three-step procedure: the soluble fraction was heated at 70°C for 20 min taking advantage from the high thermostability of the protein, successively, an affinity and a size-exclusion chromatography allowed to obtain an homogeneous sample with a yield corresponding to ∼10 mg of protein per liter of culture.

SDS/PAGE analysis showed that the purified protein migrated as a single band with an expected molecular mass of ∼10 kDa (Figure 1A, lane 2), which agrees with the theoretical (10281 Da was also taken into account the presence of the LEHHHHHH tag) as well as with the molecular mass determined by LC-SI-MS. Moreover, Stf76 contains a disulphide bridge that links cysteines 21 and 67, as obtained by LC-ESI-MS analysis (data not shown). However, the disulphide bridge seems to not affect the structure (see below) considering that CD and fluorescence spectra registered in the presence or not of a reducing agent did not show any differences (data not shown).

CD studies

Far-UV CD spectrum of Stf76 measured at 20°C was characterized by two minima, at 208 and 222 nm and one maximum at <200 nm (Figure 1B, continuous line), in agreement with a properly folded protein. The two broad minima, reasonably affected by the presence of β elements, are features indicative of the presence of both α and β secondary structure elements, as confirmed by the deconvolution of CD data performed by using the software CDPro. In addition, the protein proved to be highly stable being not significantly affected by exposure to increasing temperatures until 100°C (results not shown).

Analysis of the oligomerization state of Stf76 in solution

To assess Stf76 quaternary structure, chemical cross-linking experiments, size-exclusion chromatography coupled with a triple-angle light scattering-QELS and NMR were used.

A Stf76 tendency to oligomerize was observed in chemical cross-linking experiments performed with the homobifunctional amino-reactive reagent BS3. When Stf76 was

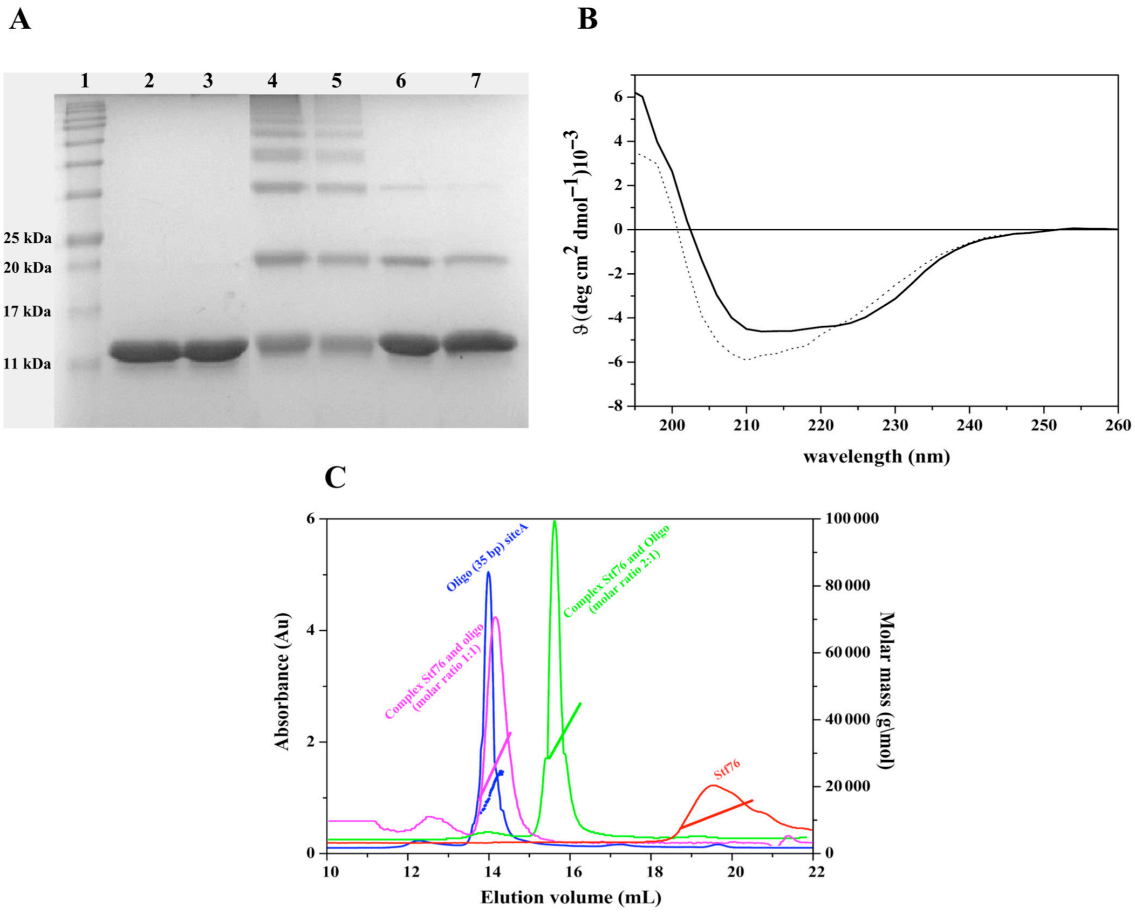


Figure 1. (**A**) Chemical cross-linking of Stf76 with BS3. Stf76 (10 μM) was incubated with and without cross-linker and analysed on SDS-PAGE. Lane 1: molecular weight markers; lane 2: recombinant Stf76; lane 3: Stf76 in the presence of DNA (10 μM); lanes 4 and 5: Stf76 in the presence of BS3 (100 μM and 200 μM, respectively); lanes 6 and 7: Stf76 in the presence of BS3 and 35 bp-site A. (**B**) Far-UV CD spectra of Stf76. CD spectra of Stf76 10 μM free (continuous line) and complexed with 5 μM of 35 bp-site A (dotted line). (**C**) Light-scattering measurements. A graph of the molecular mass and absorbance at 280 nm versus the elution volume. In red is reported Stf76; in blue 35 bp-site A; in purple the complex of Stf76 and 35 bp-site A (molar ratio 1:1), in green the complex of Stf76 and 35 bp-site A (molar ratio 2:1).

incubated with BS3 and the reaction products were separated on a denaturing gel, multiple forms of Stf76 were found (Figure 1A, lanes 4 and 5).

Static light scattering analysis showed that Stf76 has a molecular mass of about 13 kDa, which indicated that the protein is prevalently a monomer in solution (Figure 1C, in red). Dynamic light scattering (DLS) analysis conducted at concentrations ranging from 50 μM to 300 μM, indicated that Stf76 shows a concentration-dependent monomer/oligomer equilibrium in the conditions assayed. Indeed, a single peak corresponding to a monomer appeared until to 200 μM, while increasing concentrations resulted in the formation of oligomers (data not shown).

Hydrodynamic properties of Stf76 in solution were estimated at 200 μM using NMR backbone ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rate measurements (Supplementary Figure S1) in combination with DLS and structure-derived data (41,42). An estimation of the isotropic rotational correlation time, τ_c, was obtained from the R_2/R_1 ratio of the 54 residues that passed the coarse- and fine-filtering steps (43), yielding a mean ± SD value of 8.4 ± 0.7 ns. This value is slightly higher than that expected for a 10.6 kDa (taking into account ^{15}N enrichment) monomeric protein (Supplementary Figure S2A–B), as determined from a correlation of τ_c versus molecular weight of 20 known monomeric proteins studied by ^{15}N relaxation (51), suggesting that the tumbling of the protein in solution is not isotropic. According to the Stokes–Einstein–Debye equation, the obtained τ_c corresponds to an average hydrodynamic radius (r_H) of 2.09 ± 0.06 nm. This value was in perfect agreement with that measured by DLS analyses (2.05 ± 0.11 nm) and, especially, with that calculated by Hy-

dropro program (45) using the 20 models of Stf76 NMR structure here determined (2.07 ± 0.03 nm, see Supplementary Figure S2B). As a result, the experimental NMR and DLS r_H, together with that calculated from the structure, provide clear indication that the protein Stf76 is monomeric under the analyzed conditions, though exhibiting a shape not completely globular.

Stf76 is a DNA-binding protein

As already reported, the pRN1 protein ORF80 and Stf76 share 76% of sequence identity, thus they are very likely homologs (19). The DNA-binding region of ORF80 has been extensively characterized by footprinting and EMSA analyses and identified in the operator sequences (**TTAAN$_7$TTAA**) occurring twice in the region upstream of its own gene at a relative distance of 60 bases from each other. Both binding sites contain two palindromic TTAA motifs whose centers are separated by 11 bp, i.e. about one helix turn (19). In order to confirm that the homology between ORF80 and Stf76 corresponded also to a functional similarity, we analyzed the DNA-binding activity of recombinant Stf76 by EMSA experiments using DNA fragments issued from the region upstream of the *Stf76* gene. The putative binding sites for Stf76 were predicted to be two DNA regions identically spaced and structured to those of ORF80 (site A and site B, Figure 2A) and sharing 93–96% identity with them (Supplementary Figure S3). When a large 137 bp DNA region (from −120 to +17), including both putative binding sites (probe A+B, Figure 2A), was used as substrate in EMSA assay, Stf76 formed two concentration-dependent migrating complexes, indicating that the interaction presumably occurred at the two distinct previously indicated binding sites (Figure 2B). A third band, unable to enter into the gel (Figure 2B, W), probably resulted from non-specific Stf76–DNA interactions or from the formation of aggregates. The faster migrating complex (FB, Figure 2B) likely consists of a heterogeneous mixture of complexes exhibiting identical migration velocities in EMSA, but having Stf76 bound at one of the two target sites only, whereas the slower migrating complex (SB, Figure 2B) could correspond to Stf76 bound to two sites. The appearance of the two migrating complexes occurs almost simultaneously at Stf76 concentration of ≥2 μM (Figure 2B and Supplementary Figure S4) whereas at lower values the faster migrating band appears more defined compared to the slower one (Supplementary Figure S4). Differently from EMSA experiments performed with ORF80, we only observed two distinct bands and no intermediate migrating complexes have ever been detected (19). The calculated Hill coefficient from this titration is of 4.5, indicating the existence of a positive cooperativity in the Stf76 binding to the two sites. This value is in perfect agreement with that obtained for ORF80 and provides the minimum number of Stf76 molecules that bind cooperatively to the two sites (19). Binding of Stf76 to the analyzed region turned to be specific. Indeed, Stf76 was completely unable to bind to the promoter of the *f55* gene (52) used as negative control for EMSA experiments (Supplementary Figure S5). Furthermore, interaction with the 137-bp-long fragment was only slightly affected by the presence of increasing amounts of non-specific competitor DNA (Salmon Sperm DNA) at up to a 2000-fold excess (Supplementary Figure S6A). Conversely, the two complexes were displaced by increasing amount of cold specific DNA (Figure 2C). In detail, the slower migrating signal disappears at 500 molar excess, while the faster one still persists and decreases significantly at 1000 molar excess (Supplementary Figure S6B). Altogether these results demonstrate that Stf76 binds specifically to two distinct sites located in its own promoter.

Subsequently, in order to identify unambiguously the two interacting sites, we performed binding assays using double-stranded probes shortened step-wise and containing either one of the two putative binding sites. First, we tested ∼ 50 bp long probes (probe A and probe B, Figure 2A), and in both cases, only one retarded band was observed, indicating that each of the probes used did contain only one of the two putative binding sites (Figure 2D). By comparing the sequences of probe A and probe B (not shown), it was observed that only the regions including sites A and B shared a significant degree of similarity, thus supplying further indication that the interaction of Stf76 might occur at the two putative binding sites. Nevertheless, Stf76 showed differential affinity towards the two sites since a Stf76 /probe A complex is clearly evident at Stf76 concentration of ≥3 μM (Supplementary Figure S6C), whereas that constituted by Stf76 and probe B is detectable at a concentration ≥25.8 μM (Figure 2D). Furthermore, by quantification of the relative signals, Stf76 bound more efficiently (up to 5-fold) to site A than site B (Figure 2D) in the concentration range of 12.9–25.8 μM (Figure 2D, lanes 2–3 versus 2'–3'). By analyzing the sequences of site A and site B (Figure 2A), an almost perfect direct repeat (agta**ttaa**taaagtg**ttaa**t) was only detected in site A, thus providing a molecular basis for the higher affinity of Stf76 toward it. Moreover, since the slower band corresponding to the Stf76 bound either to site A or site B has already been observed at Stf76 concentration of 3 μM, the binding of Stf76 to site B seems to be positively influenced by the co-presence of site A on the same DNA fragment (Figure 2B and D).

To dissect the molecular basis of the interaction between Stf76 and its DNA sequence, we focused on the most affine site A. We defined the minimal size of the binding site that allowed specific binding. A distinct protein–DNA complex was detected only for the 35 bp probe A' (Figure 2A and E), whereas with smaller probes (25 bp and 15 bp, not shown) only aspecific aggregates were observed. Therefore, all the subsequent biochemical and structural analyses (see below) were performed with the 35 bp probe A', which retains the ability of binding Sft76 specifically.

In order to analyze the contribution of one of the two TTAA subsites of site A (**TTAA**AAAGTG**TTAA**), the second TTAA repeat was simultaneously mutagenized in four positions (probe A'*, Figure 2A). As clearly shown in Figure 2E, the binding of Stf76 was completely abolished when the mutagenized probe was used, thus confirming that at least one of the four nucleotides in the TTAA repeat is necessary for the binding to the DNA and that Stf76 interacting core sequences are indeed those initially hypothesized from comparison with the homologue ORF80 protein. Moreover, Stf76 is unable to form a stable complex with only one TTAA motif within each recognition site (Figure 2E).

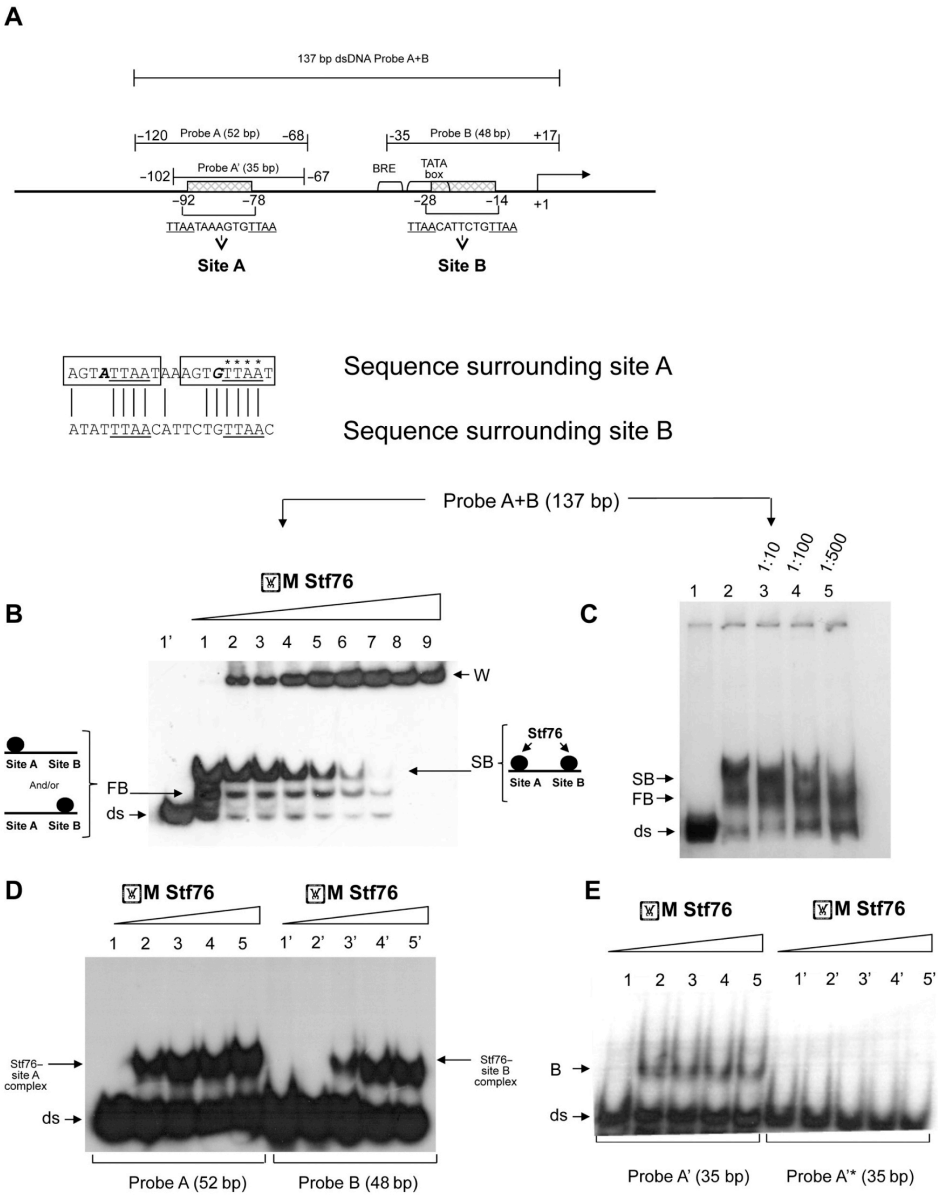


Figure 2. (**A**) Schematic representation of the DNA fragments used as probes in EMSAs. Sequences of site A and site B extending from −78 to −92 and from −14 to −28 with respect to the transcription start site (TSS), respectively, are shown. The two direct repeats found in site A are boxed and the nucleotides mutagenized in probe A'* are marked with an asterisk. (**B**–**E**) EMSA analysis of Stf76-specific binding to DNA. (**B**) Stf76 binds to a 137 bp DNA fragment (probe A+B, extending from −120/+17) that contains its own promoter in the presence of a large amount (1 μg) of non-specific competitor DNA (Salmon Sperm DNA), forming two distinct DNA-Stf76 complexes with different electrophoretic mobility. W stands for bound and aggregated DNA that remained in the wells of the gel. Binding to the labeled A+B probe (10 nM) was tested over a wide range of increasing concentration of Stf76, i.e. from 3.2, 6.4, 9.6, 12.8, 16, 19.2, 22.4, 25.8, 32.4 μM (lanes 1–9) until a complete titration of the free probe was observed. Lane 1' contains the free probe. (**C**) Lanes 3–5, labeled A+B probe incubated with purified Stf76 (3.2 μM) in the presence of increasing amounts of a specific cold competitor DNA. The ratio between labeled/unlabeled probe is shown on the top. Lane 1 and lane 2 contain the free probe and the EMSA mixture without the cold probe, respectively. A comprehensive scheme of the interaction occurring at sites A and B (slower migrating band—SB) and at either one of the two sites (faster migrating band—FB) is shown. (**D**) Comparison between binding efficiency of Stf76 towards site A and site B. Increasing amounts of Stf76 (12.8, 25.8, 40, 51.6 μM) were incubated with 5nM of labeled probe A (lanes 2–5) and probe B (lanes 2'–5'). Lanes 1 and 1' contain the free probe. (**E**) Identification of Stf76 binding determinants. Mutation of probe A'* at the four nucleotides indicated in Figure 2A abrogates Stf76 binding. Increasing amounts of Stf76 (12.8, 25.8, 40, 51.6 μM, lanes 2–5 and 2'–5') were incubated with 5 nM of labeled probe A' (lanes 1–5) or probe A'* (1'–5'). B is bound DNA.

Stf76 belongs to the 'winged helix' DNA-binding domain superfamily

Uniformly, ^{15}N- and ^{15}N/^{13}C-labeled Stf76 samples were used for NMR structural studies. [^{15}N, ^{1}H] HSQC spectra showed good cross-peak dispersion in both nitrogen and proton dimension, indicative that Stf76 adopted a well-defined structure in aqueous solution.

Nearly complete backbone assignment (H^N, N, Cα and Hα) was carried out by using data from the HNCA, HN(CO)CA, ^{15}N-edited NOESY and ^{15}N-edited TOCSY spectra. The few missing backbone assignments correspond to the residues of the N- and C-terminal ends (residues 1–2, 80–83). Side chain assignment could not be attained due to the relatively low protein concentration used to obtain Stf76 in the monomeric form. However, the side chain HN groups of Trp and Arg, as well as side chain NH_2 resonances of Gln and Asn, were detected and assigned, allowing a complete assignment of the ^{15}N-HSQC-detected cross-peaks (Figure 3A).

Secondary structure elements and their succession along the primary sequence were initially identified by CSI analysis (35), on the basis of the differences in experimental Cα and Hα chemical shift with respect to random coil values (ΔCα and ΔHα) (Figure 3B). This analysis revealed that Stf76 contained four α-helices and three β-strands, arranged in a αβααββα topology. These secondary structures were also confirmed by specific cross-peak patterns observed in the 3D ^{15}N-NOESY-HSQC (Figure 4A).

Due to the low concentration used, high-resolution structure determination with conventional NMR-derived distance restraints and *J*-couplings could not be derived. Therefore, the 3D structure of Stf76 was determined, as reported in Materials and Methods, from NMR chemical shifts by using the CS-Rosetta software (22,23).

The final ensemble of Stf76 containing the best 20 conformers (Figure 4B) is well defined in the region encompassing residues 8–68, as indicated by the root-mean-square deviations (rmsd) with respect to the mean coordinate position of the backbone and of the heavy atoms of 0.74 Å and 1.25 Å, respectively. Residues 1–7 are mostly disordered, whereas residues 72–76 showed a propensity for helix conformation, according to their deviation from random coil values (Figure 3B). PROCHECK statistics indicate that the structures have a good quality. Indeed, all backbone ϕ/ψ pairs lie mainly within the most favored and additional allowed regions of the Ramachandran plot (53), and G-factors for dihedral angle and main-chain bond lengths are positive (Table 1).

The NMR structure of Stf76 reveals a winged helix–turn–helix fold (wHTH). This fold typically contains a right-handed three helix bundle that defines the HTH motif, flanked by a two- or three-stranded antiparallel β-sheet that constitutes the wing. In the case of Stf76 (Figure 4C), the polypeptide runs from the N-terminus, through helix α1 (Leu8-Asn18), strand β1 (Cys21-Thr23), helices α2 (Leu24-Thr31) and α3 (Arg35-Arg48), and then adds strands β2 (Ile51-Phe58) and β3 (Arg61-Leu68) to complete the wing. The three-stranded version of the wHTH domain of Stf76 is often encountered in DNA-binding domains of some of the largest families of prokaryotic TFs, as well as of several eukaryotic DNA-binding domains (54). In Stf76, β2 and β3 strands are connected by a tight 1→4 turn of type I (Phe58-Arg61) and form an antiparallel β-hairpin. Three double inter-main-chain H-bonds between Tyr52-Cys67, Lys54-Glu65, Arg56-Tyr63 and Phe58-Tyr61 stabilize this hairpin (Figure 4A). The loop between helix α1 and helix α2 assumes an extended conformation and is incorporated as the third strand in the sheet via a double H-bond between backbone atoms of Tyr66 and Gln22. This arrangement brings the Cys21 close to the Cys67, confirming the presence of the disulphide bond between the two cysteines, as identified by LC-MS analysis (see above). This intramolecular disulphide bond may contribute to high Stf76 thermal stability as observed by CD studies.

^{15}N longitudinal (R_1) and transverse (R_2) relaxation constant rates were measured to support the structural data (Supplementary Figure S1). R_1 and R_2 parameters are generally constant along the whole wHTH domain as expected for a rigid structure, whereas they are outside the mean values in the N- and C-terminal regions. Therefore, these data confirm that Stf76 adopts in solution a quite rigid domain in the region 8–68 with a flexible tail in the N-terminal region and a relatively more rigid C-terminal part. Noteworthy, residues of β2–β3 loop (S59, G60) show higher values of both R_1 and R_2 compared to the mean plus one standard deviation, indicating that these regions may exhibit nanosecond motions as well as a mobility in the micro- to millisecond timescale. Higher than average R_2 values were also shown by residues of the α1 helix (L11), α1–α2 loop (G20), α3–β2 loop (K54), β2 sheet (H57) and β3 sheet (Y66), which may be a result of the presence of a low-frequency conformational or chemical exchange contribution.

Structural comparison of Stf76 wHTH with related structures

A BLASTp search using the amino acid sequence confirmed that Stf76 is closely related to plasmid regulatory proteins from *Sulfolobus* (19). However, none of these proteins has been structurally characterized so far and DNA-binding activity studies are available only for Orf80.

A search for structural homologues using DALI server (40) provided both characterized and putative winged HTH transcriptional regulators, although the levels of amino acid sequence identity were fairly low (lower than 30%). In particular, the Stf76 structure is closely related to those of SmtB/ArsR family members, including the cyanobacterial metallothionein repressor SmtB from *Synechococcus* (55), the transcriptional regulator Hlyu from *Vibrio vulnificus* (56) and the archaeal heat shock regulator Phr from *Pyrococcus furiosus* (14), with rmsd over the wHTH Cα atoms of 1.6, 1.8 and 1.9 Å, respectively. Stf76 structure is also closely related to the MarR/SlyA family members, including the organic hydroperoxide resistance regulator OhrR protein (57) from *Bacillus subtilis* in a homodimeric complex with the *ohrA* promoter (rmsd = 2.0 Å) (Figure 5).

A structure-based sequence alignment of Stf76 with these related proteins showed that the residues strictly conserved in all the sequences belong to a hydrophobic core that stabilize the fold and that are responsible for the shared struc-

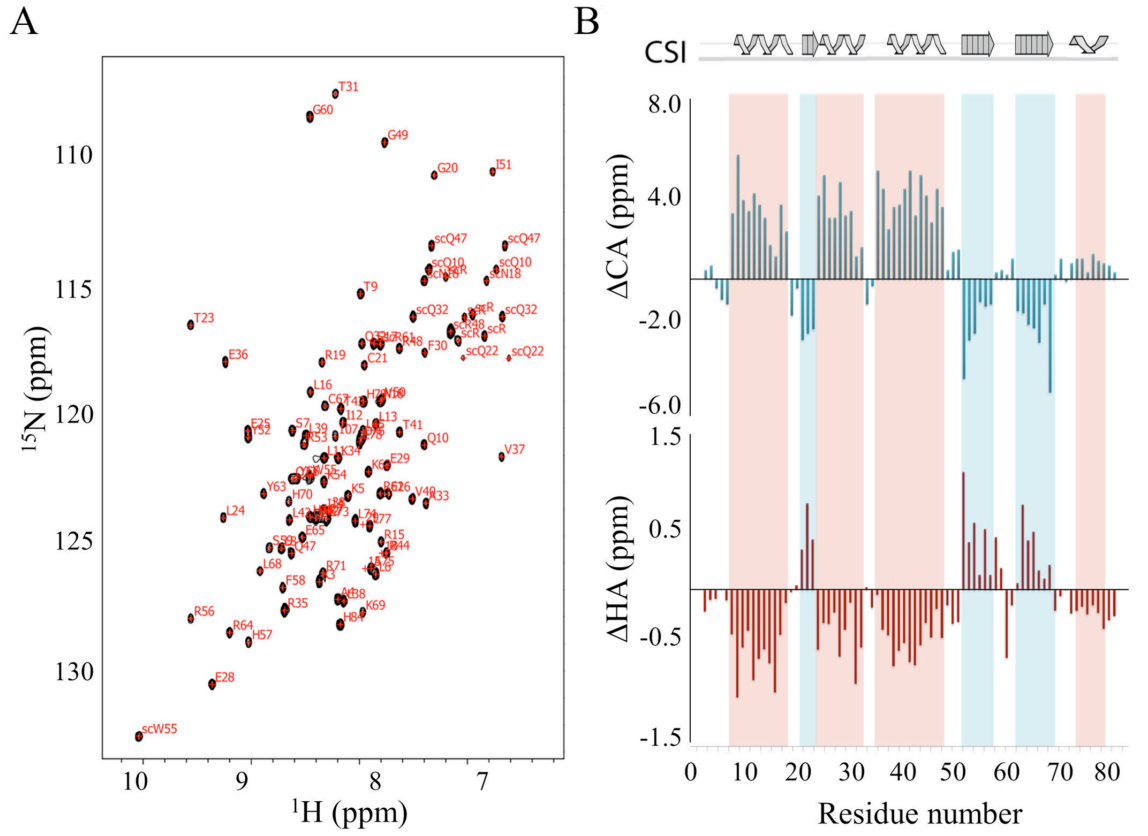


Figure 3. (**A**) 2D [^{1}H, ^{15}N] HSQC of Stf76 at 25°C and pH 5.5. Cross-peaks of the backbone amides are labeled by the one-letter amino acid code and residue numbers. Cross-peaks of side chain HNs are labeled with side chain (SC). (**B**) Chemical shift deviation from random coil values of Cα and Hα backbone atoms, obtained from CSI analysis, plotted as a function of residue number. Secondary structure elements derived from CSI are indicated above the plots. Red and blue shaded boxes indicate helical and β-strand regions, respectively.

Table 1. NMR structural statistics

Structure precision	
rmsd from mean structure (residues 5–69) (Å)	
All backbone atoms	0.99
All heavy atoms	1.69
Structure quality	
MOLPROBITY	
Clash score	0.89 ± 0.95
Poor rotamers (%)	0.05 ± 0.22
MolProbity score	0.71 ± 0.27
Residues with bad bonds (%)	0 ± 0
Residues with bad angles (%)	0 ± 0
Cβ deviations >0.25 Å	0 ± 0
PROCHECK	
G-factors phi-psi/all dihedral angles	0.32/0.49
Ramachandran plot statistics (%)	
Most favored regions	97.6
Additional allowed regions	2.4
Generously allowed regions	0
Disallowed regions	0

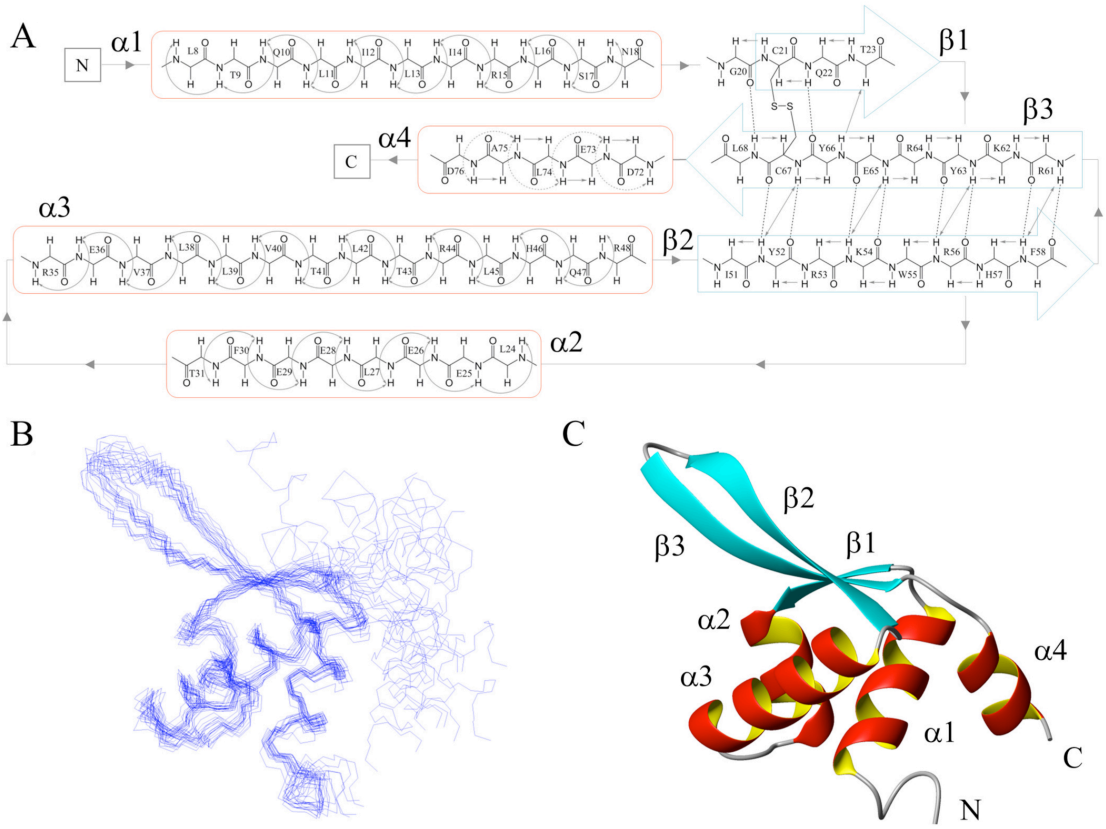


Figure 4. Topology and NMR solution structure of Stf76. (**A**) NOE network of four α-helices and the three β-strands of Stf76 as observed from the 3D NOESY-edited spectrum. NOEs are indicated by continuous arrows. In the α4 helix, weaker NOEs are indicated by dotted arrows. Hydrogen bonds are indicated by dotted lines. (**B**) Superimposition over the backbone atoms of residues 8–68 of the best 20 NMR structures obtained from CS-Rosetta (22,23). The disordered residues 1–2 and the C-terminal His-tag were omitted in the calculation (**C**) Ribbon diagram of the most representative structure of Stf76.

tural features (Figure 5A). High degree of 3D similarity is observed particularly for elements of the HTH motif, with the main differences in the length of the wing region (Figure 5B), which is known to play, together with the so-called recognition helix (α3 in Stf76), a crucial role in the contact with double-stranded DNA. According to the spatial arrangements of the helix triplets α1, α2 and α3, wHTH proteins have been classified into several structural families (58). Similarly to SmtB, Stf76 is most closely related to the toxin repressor family, characterized by having HTH angles of approximately 50° and sharing almost identical three-residue turns between α2 and α3 helices. Stf76 structure exhibits rmsd over the three helices of 1.2 Å with respect to diphtheria toxin repressor DtxR (59) (Figure 5C). Also in this case the most differences are located in the wing region.

Furthermore, compared to SmtB/ArsR and MarR/SlyA families, significant differences are observed in the flanking regions of the winged HTH domain; in particular, the N- and C-terminal helical extensions that form the dimer interface in members both of two families are absent in Stf76. This observation provides the rationale of the monomeric state of unbound Stf76 and suggests also a possible different DNA-binding mechanism with respect to these protein families.

Biochemical characterization of Stf76 DNA-binding properties

To structurally characterize the DNA-binding properties of Stf76 to site A, different analyses were performed using 35 bp-site A (see above).

The CD spectra of the protein–DNA complexes showed an increase in secondary structure elements leading to a structural reorganization of Stf76, as shown by a deeper minimum at 208 nm (Figure 1B, dashed line). The change in ellipticity increased with the amount of DNA added up to a molar ratio 2:1 (protein:DNA). Similarly, the fluorescence spectra of Stf76 registered in the presence of increasing concentration of DNA showed a decrease of emission at 355 nm until to a molar ratio 2:1 (protein:DNA) together with a slight blue-shift (Figure 6).

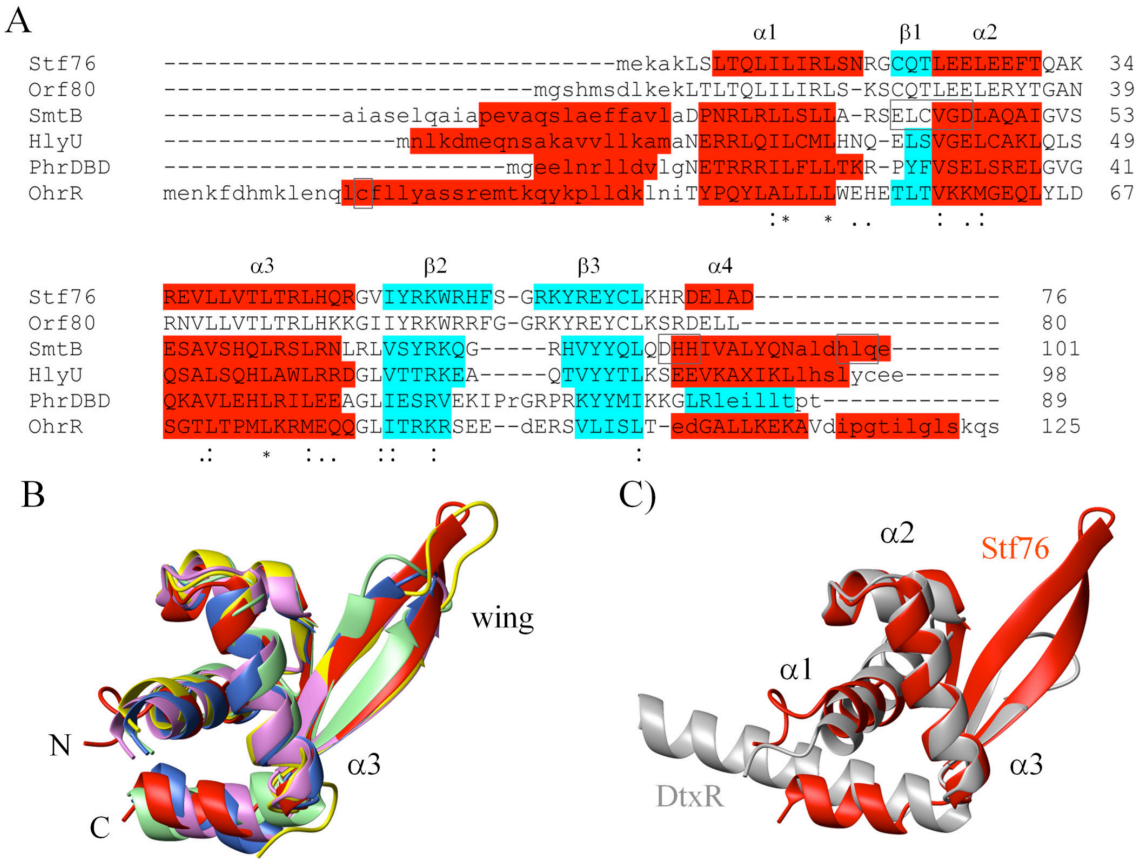


Figure 5. Sequence and structure comparison of Stf76 with related proteins. (**A**) Sequence alignment of Stf76 with its homolog ORF80 and structure-based sequence alignment of Stf76 with its closely structurally homologues SmtB, HlyU, Phr-DBD and OhrR. α-helices are shown in red, β-strands in cyan. Identical residues, strongly similar residues and somewhat similar residues are marked by an asterisk, a colon and a dot, respectively. Dissimilar residues are not marked. Structurally equivalent residues are in uppercase, structurally non-equivalent residues are in lowercase. The metal-binding sites in SmtB and OhrR C15, oxidized by organic hydroperoxides, are boxed. (**B**) Superimposed ribbon structures of the winged helix DNA-binding domains of Stf76 (red), SmtB (violet) (PDB ID 1smt chain B), HlyU (blue navy) (PDB ID 3jth chain B), Phr (yellow) (PDB ID 2p4w chain B) and Ohr (green) (PDB ID 1z9c chain D). The superposition was based on the backbone atoms of the wHTH domain excluding the α1α2 loop region and the wing loop, which show the most structural differences. (**C**) Superimposed ribbon structures over the three α-helices of the winged helix DNA-binding domains of Stf76 (red) and diphtheria toxin repressor DtxR (gray) (PDB ID 2dtr). (B) and (C) were prepared using the MolMol software (37).

Moreover, the melting curve of double-stranded oligonucleotides was monitored spectrophotometrically at 260 nm in the absence and presence of stoichiometric amounts of Stf76 (2:1). The DNA was stabilized by 10°C upon formation of the complex with Stf76, indicating that melting of the protein-bound double-stranded oligonucleotides requires more energy (data not shown).

When chemical cross-linking was performed in the presence of DNA, two predominant bands with an apparent molecular mass of 10 kDa (monomer) and 20 kDa (dimer) were produced and at the same time the oligomeric forms, observed in the absence of DNA, disappeared (Figure 1 A, lanes 6 and 7) suggesting the formation of a stable dimer only in the presence of DNA, indeed. 2:1 protein:DNA ratio is a recurring value in different techniques utilized.

Protein and annealed 35 bp oligonucleotide were mixed in a molar ratio 2:1 (protein:DNA) (Figure 1C, green) and subjected to light scattering experiments. A single peak, with a molecular weight corresponding to a 2:1 complex, was obtained with no indication of excess of either protein or DNA. Interestingly in the molecular ratio 1:1 (protein:DNA) (Figure 1C, purple), a single peak with a molecular weight corresponding to a 1:1 complex was observed indicating that the protein is also able to bind DNA as monomer. In addition, the 1:1 complex eluted at a volume similar to free DNA (Figure 1C, blue), while remarkably 2:1 complex eluted at a larger volume. The behavior of the observed complexes could be justified from the formation of more compact species (Figure 1C).

To assess the affinity of Stf76 for dsDNA, the energetics of the DNA-binding interaction between Stf76 (titrand)

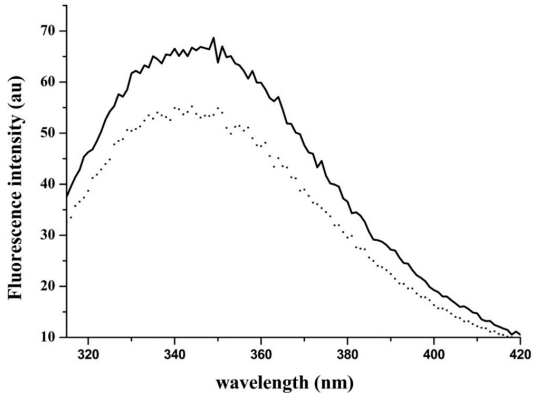


Figure 6. Emission spectra of the single tryptophan 55 of Stf76. Fluorescence spectra of Stf76 10 μM in the absence (continuous line) and in the presence of 5 μM of 35 bp-site A (dotted line).

and the selected 35 bp-site A (titrant) were determined by ITC. ITC experiment on the binding of Stf76 to 35 bp-site A showed a pattern of binding isotherm that is inconsistent with a simple 1:1 complex (Figure 7). The calorimetric isotherm is biphasic, showing endothermic heat pulses for initial injections, becoming exothermic with later additions. The ITC data showed two binding events that could be well fitted to a Sequential Binding site Model. The first binding event, with significantly higher affinity (K_1 = 40 nM), reached a plateau at about 2-fold ratio protein–DNA, while the second binding event at about 5-fold exhibited a lower binding constant (K_2) in the micromolar range. In addition, ITC measurements also showed that the tighter binding is endothermic and, interestingly, entropy driven with an unfavorable enthalpic contribution (ΔH: 10 Kcal/mol and ΔS: 70 cal/mol/deg (Figure 7).

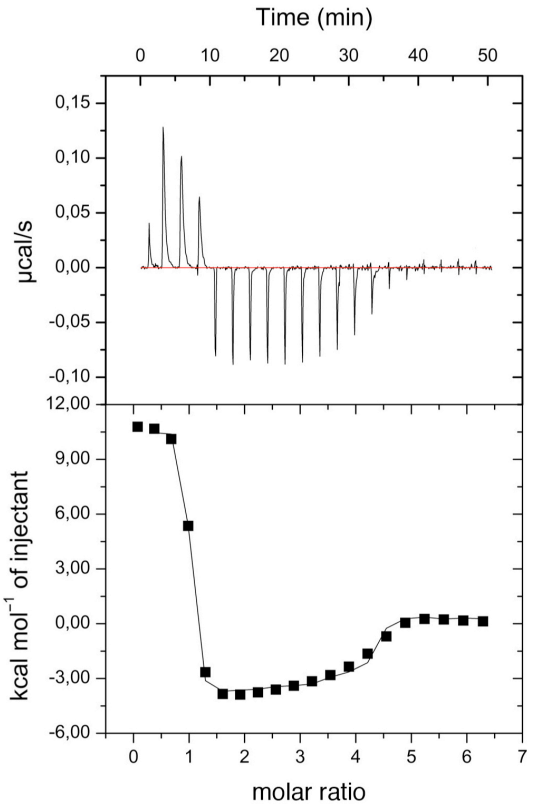


Figure 7. ITC data showing titration of 35 bp-site A (10 μM) with Stf76 (100 μM). The top and bottom sections report the raw and integrated data, respectively.

Stf76–DNA interaction by NMR CSP Analysis

To identify the DNA-binding interface of Stf76, we performed an NMR CSP analysis by acquiring 2D [^{15}N, ^{1}H] HSQC spectra of ^{15}N-labeled Stf76 in the absence and presence of the 35 bp-site A sequence (Figure 8). Upon addition of a stoichiometric amount of DNA, corresponding to 2:1 Stf76–DNA complex of ∼35 kDa, the HSQC cross-peaks of Stf76 broadened away, consistent with the formation of a large, slow tumbling complex formed between DNA and Stf76, as also indicated in the light scattering and ITC studies. At sub-stoichiometric levels of DNA, however, differential broadening and chemical shift variations were observed, providing indications on the residues of Stf76 that are direct involved in the DNA interaction (Figure 8A). Averaged combined chemical shift difference (ΔHN_{av}) and intensity changes of the amide cross-peaks upon DNA binding were plotted versus the residue number (Figure 8B) and the mapping on the structure of Stf76 of the residues showing significant changes is also reported (Figure 8C). Significant chemical shift and/or intensity changes can be observed for the residues located in the N-terminal region (Thr9, Gln10 and Leu11), in the helix α2 (Leu24, Glu25, Glu26, Glu28, Glu29), in the loop between α2 and α3 (Ala33, Lys34), in the helix α3 (Arg35, Arg44, His46, Gln47) and in the wing (Ile51, Tyr52, Arg53, Trp55, Arg56, His57, Arg58, Gly60, Lys62, Tyr63). Interestingly, all residues are mostly located on a single side of Stf76 protein, constituted by the helix α3 and the β2 and β3 strands (Figure 8C). Moreover, the distribution of the electrostatic surface potential indicates that one side of the protein has a continuous positively charged patch which includes the potential nucleic acid binding motif. On the opposite side, there are only two small patches positively charged, which do not fit with the CSPs (Supplementary Figure S7), indicating that this region is not involved in the DNA binding.

DISCUSSION

Viruses and plasmids represent attractive models to understand complex biological mechanisms, such as transcription, the cell cycle and regulation of gene expression because of their small genomes that are easy to study. Among crenarchaeal genetic elements, the pRN1 plasmid and plasmid/virus hybrid pSSVx have been biochemically and physiologically characterized (15,18,21). An open read-

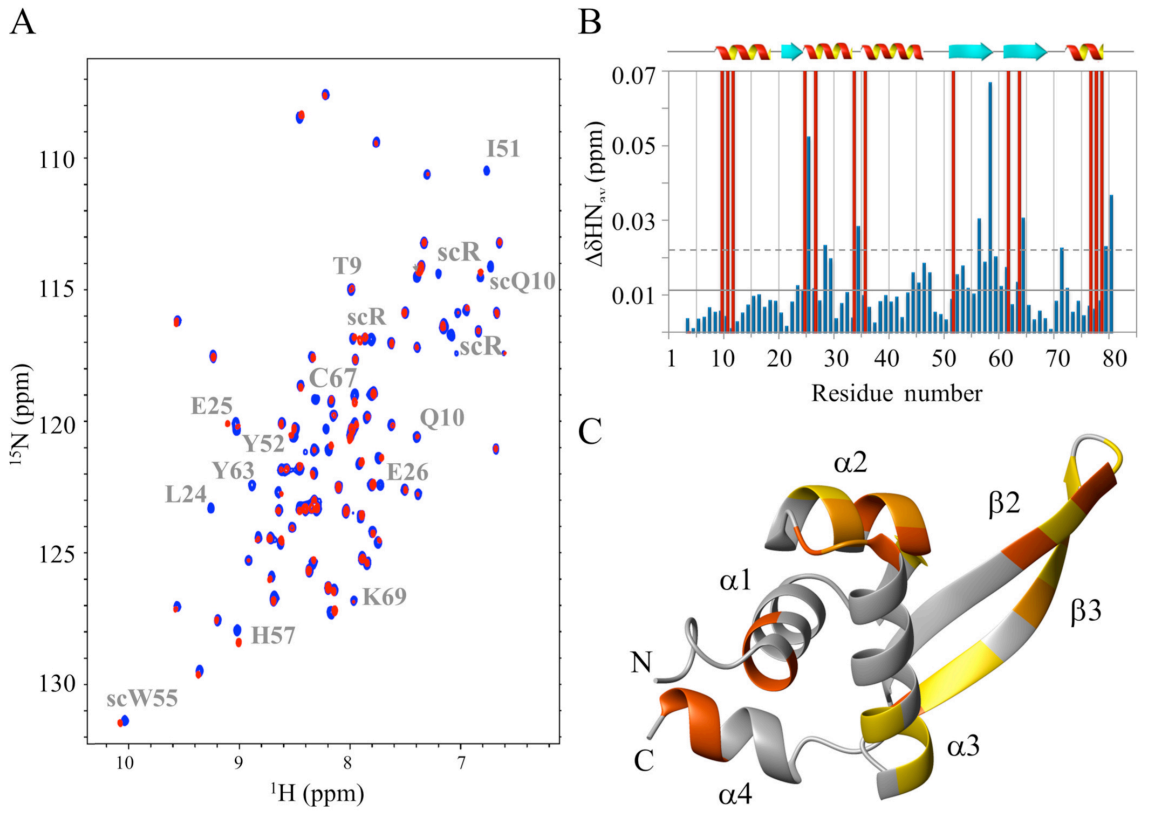


Figure 8. NMR CSP analysis of Stf76 upon DNA binding. (**A**) Superposition of a 2D [^{1}H, ^{15}N] HSQC section of Stf76 in the absence (blue) and in the presence of sub-stoichiometric amount of DNA (35 bp-site A) (red). Amide cross-peaks with significant perturbations are indicated. (**B**) Bar graphs of the average combined chemical shift differences (ΔHN_{av}) as a function of residue number. The mean value is shown as a continuous line; the mean value plus one standard deviation, as a broken line. Red bars at the maximum value in the graph represent residues with the most relative intensity reduction. The secondary structure elements are also indicated. (**C**) CSP mapping onto the representative conformer of the NMR structure of Stf76 shown as ribbon drawing. Residues for which ΔHN_{av} > mean+SD and ΔHN_{av} > mean are shown in orange and in gold, respectively. Residues with the most relative intensity reduction are shown in orange red. (C) was prepared using the software MOLMOL (37).

ing frame of pRN1, ORF80, was functionally characterized and demonstrated to be a sequence-specific double-stranded DNA-binding protein, even if no structural analyses are available so far (19,20).

In this paper, a novel protein Stf76, encoded by the hybrid plasmid–virus pSSVx and homolog of ORF80, has been functionally and structurally analysed to shed light on the molecular mechanism underpinning its interaction with DNA as well as on its biological function.

Like ORF80, Stf76 binds to its own promoter specifically at a 35-nt sequence including a 15-nt motif (TTAA-N_7-TTAA). The 15-nt sequence recurs twice in the region upstream of the *Stf76* gene, once (site B) in a partial overlapping arrangement with the B recognition element (BRE) element, thus suggesting a negative regulatory role of these cis-acting elements; in the second instance (site A), it localizes 84 nucleotides upstream of the transcription start site and therefore in a DNA region quite uncommon for a binding site of a prokaryotic transcriptional regulator. The two Stf76 binding sites are not equivalent since the recombinant protein interacts more strongly with the distal (site A) than with the proximal site (site B) (Figure 2D). The differential affinity can be explained at molecular level by the fact that while the former contains two almost perfect direct repeats (AGTATTAATAAAGTGTTAAT), site B lacks of this kind of symmetry (Figure 2A). After mutagenesis of one of the two TTAA sub-sites, the binding ability of Stf76 is completely abrogated, as shown by EMSA assay, further supporting the hypothesis that both sub-sites (5'-TTAA-3') are required for the formation of stable complexes on the DNA. Our EMSA data indicate not only that the TTAA core of the motif is essential for the binding, but also that the flanking nucleotides might play a relevant role in the stabilization. Overall, our data were in good agreement with those observed in the atomic force microscopy (AFM) study of the ORF80 interaction with dsDNA in which two ORF80 molecules were visible to bind one TTAA-N7-TTAA motif. Single complexes, preferentially formed by interaction of ORF80 to one of the two sites, were also clearly distin-

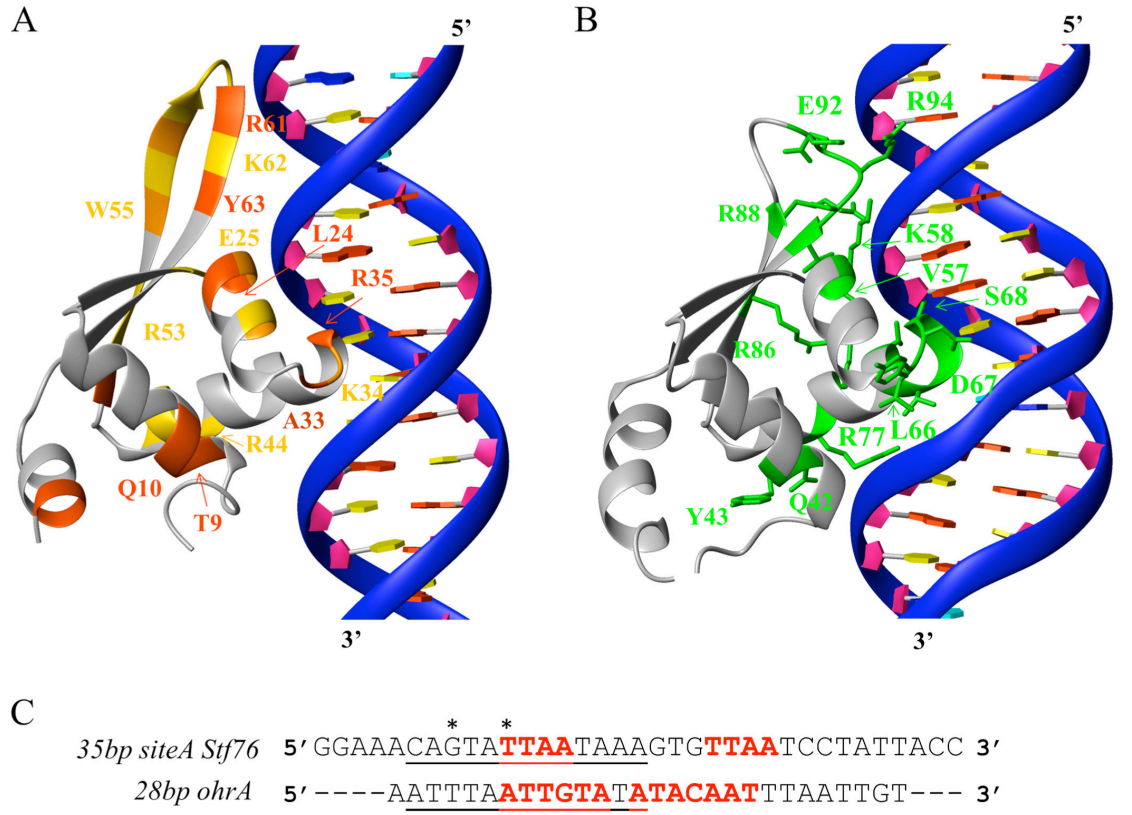


Figure 9. Model of the Stf76–DNA interaction. (**A**) One Stf76 monomer was positioned on a single sub-site of the 35 bp-site A on the basis of similarity with OhrR (ref). The backbone atoms of the Stf76 NMR structure were superimposed onto those of the crystal structure of the OhrR–*ohrA* DNA complex, reported in (**B**) (PDB code 1Z9C). In (A), the DNA of OhrR was replaced by a straight B form of the DNA, modeled by using 3D-DART (68). The backbone atoms of *Stf76* DNA were superimposed onto those of *ohrA* DNA, aligning the DNA sequences as reported in (**C**). The underlined bases are those represented in (A) and (B) panels. The two TTAA motifs in *Stf76* and the inverted repeats in *ohrA* are highlighted in bold red. Two bases, possibly involved in specific interactions, are indicated by an asterisk. In (A), DNA-binding elements of Stf76 are colored as in Figure 8C; in (B), those of OhrR are shown in green. Homologue residues in Stf76, that in OhrR interact with DNA, are labeled. Residues of the proteins that are not involved in DNA interaction are shown as a gray ribbon. Backbone atoms of the double-stranded DNAs are represented as ribbon in blue. A, T, G and C base rings and sugar rings are indicated as plates in orange red, yellow, green, cyan and pink, respectively.

guishable. Interestingly enough, the most affine ORF80 site corresponds to Stf76 site A.

The NMR structure of Stf76, here determined, indicates that Stf76 belongs to the wHTH superfamily (Figure 4), which includes the majority of *Sulfolobus* TFs. Since ORF80 and Stf76 share 76% of sequence identity, the sequence homology likely corresponds also to a structural homology (Figure 5A). ORF80 was initially predicted to bear a leucine zipper motif on the basis of the position of five completely conserved leucine (19). Subsequently, it was hypothesized to be a winged helix protein, as it exhibits a weak similarity in amino acid sequence and (predicted) secondary structure to several sequence-specific DNA-binding proteins with the wHTH fold (20). Our data confirm indirectly that also ORF80 is a novel member of the wHTH superfamily. Four of the five conserved leucine (Leu13, Leu16, Leu24 and Leu27 in Stf76) belong to the hydrophobic core that stabilize the fold, thus having a role different from that initially predicted (19).

Stf76 displays the three stranded version of the wHTH domain (54) in which the tight three-helical core is followed by a C-terminal hairpin (the wing) that forms a β-sheet with a short β-strand in the loop between α1 and α2 helices. The relaxation data confirm that region 8–68 is substantially rigid with some degree of flexibility in β2–β3 loop, whereas the N-terminus is largely flexible and the C-terminus more rigid with helical propensity (Supplementary Figure S1).

A DALI analysis showed that the Stf76 fold is structurally related to members of bacterial and archaeal SmtB/ArsR and MarR/SlyA families. SmtB/ArsR family is a class of transcription regulator involved in stress response to heavy metal ions. The structural homolog of Stf76, SmtB, is one of the best characterized members of this family (60–62). It is a homodimer that functions by binding to its operator pro-

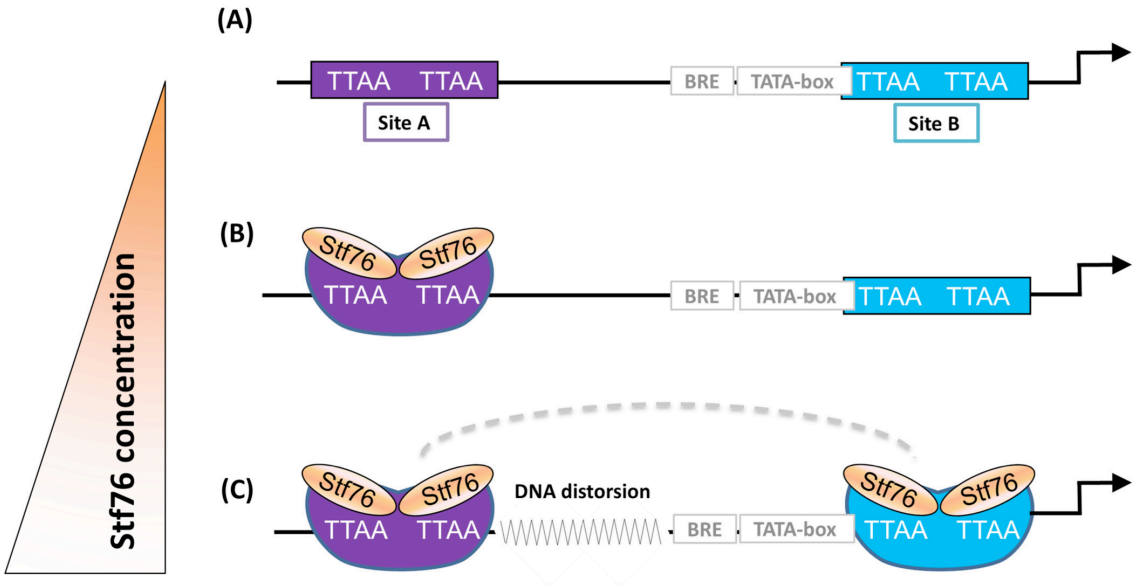


Figure 10. A simplified model for the Stf76 interaction to its own promoter. The two sites are saturated in a concentration-dependent manner according to the differential affinity of Sft76 toward them. (**A**) The promoter region containing the two binding sites, A and B, for Stf76. (**B**) At low concentrations, Stf76 binds only to the most affine site A. (**C**) This binding induces DNA distortion (indicated by shape modification of the binding site) that presumably extends to the adjacent regions in the promoter (as shown by wavy lines). This could lead to positive cooperative binding to site B at higher concentrations (indicated by a dotted line).

moter (two inverted repeats), and dissociating in the presence of Zn^{2+}. The comparison of the sequence of Stf76 with SmtB shows that the metal-binding site residues of SmtB are substituted in Stf76 with residues not able to chelate any metal ions (Figure 5A), suggesting a different function for Stf76. However, also HlyU and Phr, the other two ArsR/SmtB family members closely structurally related to Stf76, do not respond to heavy metals for the absence of metal-binding sites.

MarR/SlyA family is a class of transcription regulators involved in the development of antibiotic resistance. The mechanism by which its member OhrR responds to organic hydroperoxides is based on the oxidation of the Cys15, localized in the N-terminal helical extension of the wHTH domain, to Cys-sulphenic acid. This residue, as shown in Figure 5A, is absent in Stf76, indicating also in this case an unrelated function between Stf76 and OhrR.

As expected, the wHTH domain of Stf76 is dominated by positive charges, but it shows some atypical properties with respect to canonical winged HTH proteins (Supplementary Figure S6C, D). Indeed, typical wHTH have a basic surface mainly located on the recognition helix (63). In contrast, Stf76 exhibits clustering of basic residues on the wing (R53, K54, R56, R61, K62, R64) but displays only few basic residues on the recognition helix (R34, R44, R48). Interestingly, a similar surface charge distribution is observed in the structurally related proteins (Figure 5A), suggesting a similar binding interface.

wHTH domains are often combined in the same protein with other domains or enriched with other secondary elements which contribute to protein dimerization (54). Furthermore, wHTH proteins generally bind as dimers to palindromic or direct repeat DNA sequences in which each monomer recognizes a half site. Stf76 presents further peculiar features with respect to the other members of the wHTH superfamily because it lacks the additional elements involved in the dimerization and it is a monomer in solution. Nevertheless, chemical-cross-linking, light scattering, CD, ITC and spectrofluorimetric experiments revealed that interaction between two monomers to form stable dimers occurs, but only in the presence of DNA (64). In particular, CD and spectrofluorimetric titrations indicate the formation of a 2:1 protein:DNA complex, when Stf76 interacts with the fragment 35 bp-site A.

Far-UV CD spectra of free Stf76 showed a properly folded protein with an α/β fingerprint. In the presence of DNA, CD analyses indicate that Stf76 undergoes to an increase in helical content upon binding to the target DNA. As observed by NMR CSI analyses and ^{15}N relaxation studies (Figure 3B and Supplementary Figure S1), the three-helical core is well defined and more rigid, whereas the C-terminal helix, α4, shows the highest flexibility in the free protein. Therefore, upon binding to DNA, the disorder to order transition could involve the C-terminal helix of Stf76, which is indeed significantly affected by protein–DNA interaction (Figure 8).

In addition, spectrofluorimetric spectra showed that the only Trp present in the sequence (Trp55) results to be exposed in the free protein (λ_{max} emiss = 355 nm), in agreement with the NMR structure. In the presence of increasing DNA amounts Trp55 moves to a less polar environment as indicated by a slight blue-shift (λ_{max} emiss = 350 nm) with a concomitant decrease in emission due to quenching for the formation of DNA–protein complex. Accordingly, the Trp55 is located in the β2 strand of the basic C-terminal hairpin that establishes contacts with DNA as indicated by NMR data. In this context, the comparative DNA-binding model (see below) suggests that the interaction occurs with the minor groove.

Direct protein–DNA recognition interactions are usually characterized by favorable ΔH and favorable ΔS deriving from water molecules release, while those that strongly distort the DNA have net unfavorable ΔH, primarily associated with the base pair destacking (65). Our ITC data show a favorable entropic contribution and an unfavorable enthalpic change in the formation of the complex Stf76–DNA. Significant DNA distortions might be associated with the formation of stable Stf76 dimer deriving from a forced functional interaction of two monomers when bound to DNA. Accordingly light scattering data indicate the formation of a 2:1 complex, being more compact than the 1:1 complex due to a considerable DNA distortion.

Two potential pathways for DNA binding of protein dimers have been described: (i) a protein pre-exists in solution as a dimer and binds to DNA in this oligomeric form; (ii) the protein dimerizes upon sequential binding of two monomers to the DNA. Based on our results, the second pathway better describes the binding mode of Stf76 and would suggest that the shift of the monomer–dimer equilibrium dependent on the DNA interaction represents a level of regulation of Stf76 activity, as already described for eukaryotic TFs (66,67).

The close structural similarity among Stf76 and the DNA-binding domain of ArsR/SmtB and MarR/SlyA family members suggests a likewise pattern of DNA–protein interaction. Using NMR CSP studies of Stf76 in the presence of the 35 bp fragment of site A, we defined a DNA-binding surface constituted by the N-terminal portion of α1, α2 and α3 helices, the α2α3 loop and the β-hairpin of the wing region that well superimpose on the positively charged region (Supplementary Figure S4). Moreover, also some residues of the C-terminal α4 helix seem to be affected upon DNA binding, even if it seems to be not part of the surface directly involved in the DNA–protein interaction. Possibly, these residues could be involved in protein–protein interactions stabilizing the dimer, but more detailed structural data will be needed to clarify this point.

To date, structures in complex with DNA are still unavailable for members of the ArsR/SmtB family, whereas the only complex structure available for the MarR/SlyA family is represented by the transcriptional regulator BsOhrR bound to the 28-bp duplex *ohrA* operator. Since Stf76 exhibits a high degree of structural similarity to the OhrR wHTH domain, as shown by the DALI analysis, we used the structure of the OhrR-wHTH in complex with its DNA target as a template to generate a model of a single Stf76 monomer binding to DNA. We first superimposed the backbone atoms of Stf76 wHTH domain onto those of OhrR, retaining the Stf76 binding site as the *ohrA* target DNA. Although a bent DNA model could be also plausible, in the absence of molecular experimental data we used a straight B form DNA of the 35 bp-site A. Thus, the DNA backbone atoms of *Stf76* were superimposed onto those of *ohrA* DNA, positioning the TTAA motif in the major groove, where OhrR and, hence, Stf76 establish specific contacts (Figure 9). In the modeled structure, there is a wHTH canonical mode of DNA recognition, in which the α3 helix is presented to the major groove specifically contacting DNA base pairs. The wing and the N-terminal portion also make contacts with the minor groove of the DNA. Remarkably, the hypothesized DNA-binding surface of Stf76, as determined by NMR CSP studies, is well consistent with DNA recognition mode of OhrR. In particular, DNA binding conserved residues of Stf76 on the relative position of OhrR are: Leu24 (Val57) and Glu25 (Lys58) in the α2 helix, Lys32 (Asp67) in the loop α2–α3, Arg44 (Arg77) in the recognition α3 helix, Arg53 (Arg86) in the β2 strand, Arg61 (Glu93) and Lys62 (Arg94) in the β3 strand. Most of the differences are observed in the α3 helix that could be justified by a different site-specific DNA recognition between OhrR and Stf76. In particular, Arg35 in the α3 helix of Stf76, replaced with Ser68 in OhrR, could make a specific interaction with the first thymine of the TTAA motif. In the wing of OhrR Arg94, which establishes a specific contact with the minor groove, is replaced in Stf76 with the Lys62. The latter points outside the DNA-binding site in the model and is less affected by the binding than the adjacent Arg61 and Tyr63 (see Figure 8b), letting hypothesize that one of these two residues is likely involved in DNA recognition. In particular, a specific interaction could possibly be established with the guanine, just preceding the TTAA site, analogously to what made between Arg 94 and the corresponding thymine in the OhrR–*ohrA* complex (Figure 9). Accordingly, the long wing of Stf76 contributes to an elongated shape of the protein, in line with the hydrodynamic analysis, and accounts for the requirement to contact extra nucleotides flanking the TTAA sub-site for the stabilization of specific DNA binding.

Our previous northern blot analysis indicated that the *Stf76* gene undergoes to a dual transcription regulation entailing the occurrence of a considerable activation and a complete abrogation of its own expression depending on the pSSVx life cycle and host growth (16). Indeed, mRNA level increases progressively reaching its highest intracellular concentration at the end of the exponential growth before suddenly dropping upon the onset of a strong induction of pSSVx replication. Since the binding sites of Stf76 are located within its own promoter, it is conceivable that Stf76 could play a regulative role of its own expression.

Although our data do not indicate whether Stf76 exerts its physiological function through interaction with one or both sequences (site A and site B), binding most likely starts from the most affine primary site A and subsequently extends to site B in a concentration-dependent manner. Our EMSA data indicate a positive cooperativity in the binding of Stf76 to its interacting sites. Because of the remarkable distance between site A and site B, at this stage no exact mechanism for the cooperativity observed, can be pro-

posed. Based on the DNA distortions highlighted during ITC and light scattering experiments, one hypothesis is that Stf76 interaction to site A might induce a conformational change that favors the binding of Sft76 to site B. Since 60 bp corresponds roughly to six complete double-helix turns, the cooperativity observed might rely on the two binding sites facing the same side of the helix. A hypothetical model obtained by combining all the data here reported is presented in Figure 10. Interestingly, the putative binding sites of Stf76 homologues are well conserved both in the general structure and in the sequence. In particular, the retention of the distance between the centres of the two TTAA motifs within one binding site as well as the relative spacing of 60 bp between the two TTAAN7TTAA binding sites indicate that the regulation activity of Stf76 and of its homologues strictly depends on these specific features of the DNA-binding region.

Homologues of Stf76 are widespread among *Sulfolobus* cryptic and conjugative plasmids, suggesting a fundamental function for these TFs in the maintenance and/or in the life cycle of the plasmids in the host cells. It is tempting to speculate that Stf76 exerts its regulation activity not only by targeting its own promoter but also the regulative sequences of host genes involved in the replication and/or maintenance of the appropriate plasmids copy number (19).

Stf76 represents the first example of a viral protein belonging to the wHTH family which has been deeply characterized by means of a multidisciplinary approach including biochemical, molecular biology and NMR techniques, while other wHTH viral proteins are known only at 3D level by crystallographic techniques.

The results of this study show that Stf76 is a novel representative of the wHTH family. Compared to other members of the same family Stf76 presents several peculiar features: (i) it is a monomer in solution, (ii) it does not require additional secondary structure elements to exert its physiological function, (iii) residues that are usually responsible for effector binding in other prokaryotic wHTH TFs characterized so far are not present, (iv) the location as well as the relative distance of the two recognition motifs (sites A and B) suggests that the DNA recognition mechanism and the mode of action are unusual and potentially novel.

ACCESSION NUMBERS

The structure and chemical shifts of Stf76 have been deposited in the PDB protein data bank (http://www.pdb.org) and the BMRB database (http://www.bmrb.wisc.edu) under the accession numbers 2MLG and 19821, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Leopoldo Zona for providing excellent technical assistance.

FUNDING

Italian MIUR (PRIN 2010 2010M2JARJ_002 and FIRB RBAP114AMK 006); Programma MERIT (RBNE08HWLZ_014) from Ministero dell' Istruzione, dell'Università e della Ricerca; "Programma F.A.R.O, IV° tornata -Università Federico II di Napoli and "Compagnia di San Paolo" (Naples Laboratory). (PON_1078) "Identificazione di biomarcatori e sviluppo di metodi diagnostici e terapeutici nel campo dell'oncologia e della biologia vascolare".

Conflict of interest statement. None declared.

REFERENCES

1. Contursi,P., D'Ambrosio,K., Pirone,L., Pedone,E., Aucelli,T., She,Q., De Simone,G. and Bartolucci,S. (2011) C68 from the Sulfolobus islandicus plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. *Biochem. J.*, **435**, 157–166.
2. Prangishvili,D. (2013) The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.*, **67**, 565–585.
3. Krupovic,M., White,M.F., Forterre,P. and Prangishvili,D. (2012) Postcards from the edge: structural genomics of archaeal viruses. *Adv. Virus Res.*, **82**, 33–62.
4. Prangishvili,D., Garrett,R.A. and Koonin,E.V. (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res.*, **117**, 52–67.
5. Peng,X., Garrett,R.A. and She,Q.X. (2012) Archaeal viruses-novel, diverse and enigmatic. *Sci. China Life Sci.*, **55**, 422–433.
6. Guilliere,F., Peixeiro,N., Kessler,A., Raynal,B., Desnoues,N., Keller,J., Delepierre,M., Prangishvili,D., Sezonov,G. and Guijarro,J.I. (2009) Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. *J. Biol. Chem.*, **284**, 22222–22237.
7. Kessler,A., Brinkman,A.B., van der Oost,J. and Prangishvili,D. (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon sulfolobus. *J. Bacteriol.*, **186**, 7745–7753.
8. Kessler,A., Sezonov,G., Guijarro,J.I., Desnoues,N., Rose,T., Delepierre,M., Bell,S.D. and Prangishvili,D. (2006) A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. *Nucleic Acids Res.*, **34**, 4837–4845.
9. Schleper,C., Kubo,K. and Zillig,W. (1992) The particle Ssv1 from the extremely thermophilic archaeon sulfolobus is a virus - demonstration of infectivity and of transfection with viral-DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 7645–7649.
10. Contursi,P., Jensen,S., Aucelli,T., Rossi,M., Bartolucci,S. and She,Q.X. (2006) Characterization of the Sulfolobus host-SSV2 virus interaction. *Extremophiles*, **10**, 615–627.
11. Contursi,P., Fusco,S., Limauro,D. and Fiorentino,G. (2013) Host and viral transcriptional regulators in Sulfolobus: an overview. *Extremophiles*, **17**, 881–895.
12. Leonard,P.M., Smits,S.H., Sedelnikova,S.E., Brinkman,A.B., de Vos,W.M., van der Oost,J., Rice,D.W. and Rafferty,J.B. (2001) Crystal structure of the Lrp-like transcriptional regulator from the archaeon Pyrococcus furiosus. *Embo. J.*, **20**, 990–997.
13. Shinkai,A., Sekine,S., Urushibata,A., Terada,T., Shirouzu,M. and Yokoyama,S. (2007) The putative DNA-binding protein Sto12a from the thermoacidophilic archaeon Sulfolobus tokodaii contains intrachain and interchain disulfide bonds. *J. Mol. Biol.*, **372**, 1293–1304.
14. Liu,W., Vierke,G., Wenke,A.K., Thomm,M. and Ladenstein,R. (2007) Crystal structure of the archaeal heat shock regulator from Pyrococcus furiosus: a molecular chimera representing eukaryal and bacterial features. *J. Mol. Biol.*, **369**, 474–488.
15. Arnold,H.P., She,Q., Phan,H., Stedman,K., Prangishvili,D., Holz,I., Kristjansson,J.K., Garrett,R. and Zillig,W. (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon Sulfolobus is a hybrid between a plasmid and a virus. *Mol. Microbiol.*, **34**, 217–226.
16. Contursi,P., Cannio,R., Prato,S., She,Q., Rossi,M. and Bartolucci,S. (2007) Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. *J. Bacteriol.*, **189**, 6339–6350.

17. Contursi,P., Cannio,R. and She,Q.X. (2010) Transcription termination in the plasmid/virus hybrid pSSVx from Sulfolobus islandicus. *Extremophiles*, **14**, 453–463.
18. Lipps,G. (2006) Plasmids and viruses of the thermoacidophilic crenarchaeote Sulfolobus. *Extremophiles*, **10**, 17–28.
19. Lipps,G., Ibanez,P., Stroessenreuther,T., Hekimian,K. and Krauss,G. (2001) The protein ORF80 from the acidophilic and thermophilic archaeon Sulfolobus islandicus binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. *Nucleic Acids Res.*, **29**, 4973–4982.
20. Lysetska,M., Zettl,H., Oka,I., Lipps,G., Krauss,G. and Krausch,G. (2005) Site-specific binding of the 9.5 kilodalton DNA-binding protein ORF80 visualized by atomic force microscopy. *Biomacromolecules*, **6**, 1252–1257.
21. Berkner,S. and Lipps,G. (2007) Characterization of the transcriptional activity of the cryptic plasmid pRN1 from Sulfolobus islandicus REN1H1 and regulation of its replication operon. *J. Bacteriol.*, **189**, 1711–1721.
22. Shen,Y., Lange,O., Delaglio,F., Rossi,P., Aramini,J.M., Liu,G., Eletsky,A., Wu,Y., Singarapu,K.K., Lemak,A. *et al.* (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 4685–4690.
23. Shen,Y., Vernon,R., Baker,D. and Bax,A. (2009) De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR*, **43**, 63–78.
24. Aucelli,T., Contursi,P., Girfoglio,M., Rossi,M. and Cannio,R. (2006) A spreadable, non-integrative and high copy number shuttle vector for Sulfolobus solfataricus based on the genetic element pSSVx from Sulfolobus islandicus. *Nucleic Acids Res.*, **34**, e114.
25. Limauro,D., Saviano,M., Galdi,I., Rossi,M., Bartolucci,S. and Pedone,E. (2009) Sulfolobus solfataricus protein disulphide oxidoreductase: insight into the roles of its redox sites. *Protein Eng. Des. Sel.*, **22**, 19–26.
26. Mercurio,F.A., Marasco,D., Pirone,L., Pedone,E.M., Pellecchia,M. and Leone,M. (2012) Solution structure of the first Sam domain of Odin and binding studies with the EphA2 receptor. *Biochemistry*, **51**, 2136–2145.
27. Kay,L.E., Ikura,M., Tschudin,R. and Bax,A. (2011) Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J. Magn. Reson.*, **213**, 423–441.
28. Farmer,B.T. II, Venters,R.A., Spicer,L.D., Wittekind,M.G. and Muller,L. (1992) A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules. *J. Biomol. NMR*, **2**, 195–202.
29. Grzesiek,S. and Bax,A. (1992) Improved 3d triple-resonance NMR techniques applied to a 31-Kda protein. *J. Magn. Reson.*, **96**, 432–440.
30. Bax,A. and Ikura,M. (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J. Biomol. NMR*, **1**, 99–104.
31. Marion,D., Driscoll,P.C., Kay,L.E., Wingfield,P.T., Bax,A., Gronenborn,A.M. and Clore,G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*, **28**, 6150–6156.
32. Marion,D., Kay,L.E., Sparks,S.W., Torchia,D.A. and Bax,A. (1989) 3-Dimensional Heteronuclear Nmr of N-15-Labeled Proteins. *J. Am. Chem. Soc.*, **111**, 1515–1517.
33. Zuiderweg,E.R. and Fesik,S.W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry*, **28**, 2387–2391.
34. Keller, (ed.) (2004) *The Computer Aided Resonance Assignment Tutorial*. CANTINA, Germany.
35. Wishart,D.S. and Case,D.A. (2001) Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.*, **338**, 3–34.
36. Shen,Y., Delaglio,F., Cornilescu,G. and Bax,A. (2009) TALOS plus: A hybrid method for predicting protein torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
37. Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.
38. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
39. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. 3rd, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–383.
40. Hasegawa,H. and Holm,L. (2009) Advances and pitfalls of protein structural alignment. *Curr. Opin. Struc. Biol.*, **19**, 341–348.
41. Malgieri,G., Russo,L., Esposito,S., Baglivo,I., Zaccaro,L., Pedone,E.M., Di Blasio,B., Isernia,C., Pedone,P.V. and Fattorusso,R. (2007) The prokaryotic Cys2His2 zinc-finger adopts a novel fold as revealed by the NMR structure of Agrobacterium tumefaciens Ros DNA-binding domain. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 17341–17346.
42. Farina,B., Pirone,L., Russo,L., Viparelli,F., Doti,N., Pedone,C., Pedone,E.M. and Fattorusso,R. (2010) NMR backbone dynamics studies of human PED/PEA-15 outline protein functional sites. *FEBS J.*, **277**, 4229–4240.
43. Pawley,N.H., Wang,C.Y., Koide,S. and Nicholson,L.K. (2001) An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of N-15 relaxation parameters. *J. Biomol. NMR*, **20**, 149–165.
44. Kay,L.E., Torchia,D.A. and Bax,A. (1989) Backbone dynamics of proteins as studied by N-15 inverse detected heteronuclear NMR-spectroscopy - application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
45. Ortega,A., Amoros,D. and Garcia de la Torre,J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.*, **101**, 892–898.
46. Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J. and Gronenborn,A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA-binding domain of Gata-1. *Science*, **261**, 438–446.
47. Foster,M.P., Wuttke,D.S., Clemens,K.R., Jahnke,W., Radhakrishnan,I., Tennant,L., Reymond,M., Chung,J. and Wright,P.E. (1998) Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *J. Biomol. NMR*, **12**, 51–71.
48. Grzesiek,S., Dobeli,H., Gentz,R., Garotta,G., Labhardt,A.M. and Bax,A. (1992) 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry*, **31**, 8180–8190.
49. Garrett,D.S., Seok,Y.J., Peterkofsky,A., Clore,G.M. and Gronenborn,A.M. (1997) Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system. *Biochemistry*, **36**, 4393–4398.
50. Grzesiek,S., Bax,A., Clore,G.M., Gronenborn,A.M., Hu,J.S., Kaufman,J., Palmer,I., Stahl,S.J. and Wingfield,P.T. (1996) The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat. Struct. Biol.*, **3**, 340–345.
51. Rossi,P., Swapna,G.V., Huang,Y.J., Aramini,J.M., Anklin,C., Conover,K., Hamilton,K., Xiao,R., Acton,T.B., Ertekin,A. *et al.* (2010) A microscale protein NMR sample screening pipeline. *J. Biomol. NMR*, **46**, 11–22.
52. Fusco,S., She,Q., Bartolucci,S. and Contursi,P. (2013) T(lys), a newly identified Sulfolobus spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. *J. Virol.*, **87**, 5926–5936.
53. Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, **7**, 95–99.
54. Aravind,L., Anantharaman,V., Balaji,S., Babu,M.M. and Iyer,L.M. (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.
55. Cook,W.J., Kar,S.R., Taylor,K.B. and Hall,L.M. (1998) Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. *J. Mol. Biol.*, **275**, 337–346.

56. Nishi,K., Lee,H.J., Park,S.Y., Bae,S.J., Lee,S.E., Adams,P.D., Rhee,J.H. and Kim,J.S. (2010) Crystal structure of the transcriptional activator HlyU from Vibrio vulnificus CMCP6. *FEBS Lett.*, **584**, 1097–1102.
57. Hong,M., Fuangthong,M., Helmann,J.D. and Brennan,R.G. (2005) Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol. Cell.*, **20**, 131–141.
58. Wintjens,R. and Rooman,M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J. Mol. Biol.*, **262**, 294–313.
59. Qiu,X.Y., Pohl,E., Holmes,R.K. and Hol,W.G.J. (1996) High-resolution structure of the diphtheria toxin repressor complexed with cobalt and manganese reveals an SH3-like third domain and suggests a possible role of phosphate as co-corepressor. *Biochemistry*, **35**, 12292–12302.
60. Eicken,C., Pennella,M.A., Chen,X., Koshlap,K.M., VanZile,M.L., Sacchettini,J.C. and Giedroc,D.P. (2003) A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J. Mol. Biol.*, **333**, 683–695.
61. Bartolucci,S., Contursi,P., Fiorentino,G., Limauro,D. and Pedone,E. (2013) Responding to toxic compounds: a genomic and functional overview of Archaea. *Front. Biosci. (Landmark Ed)*, **18**, 165–189.
62. Del Giudice,I., Limauro,D., Pedone,E., Bartolucci,S. and Fiorentino,G. (2013) A novel arsenate reductase from the bacterium Thermus thermophilus HB27: Its role in arsenic detoxification. *Biochim. Biophys. Acta*, **1834**, 2071–2079.
63. Gajiwala,K.S. and Burley,S.K. (2000) Winged helix proteins. *Curr. Opin. Struc. Biol.*, **10**, 110–116.
64. Wadsworth,R.I. and White,M.F. (2001) Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus. *Nucleic Acids Res.*, **29**, 914–920.
65. Jen-Jacobson,L., Engler,L.E. and Jacobson,L.A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure*, **8**, 1015–1023.
66. Park,C., Campbell,J.L. and Goddard,W.A. (1996) Can the monomer of the leucine zipper proteins recognize the dimer binding site without dimerization? *J. Am. Chem. Soc.*, **118**, 4235–4239.
67. Chin,J.W., Kohler,J.J., Schneider,T.L. and Schepartz,A. (1999) Gene regulation: Protein escorts to the transcription ball. *Curr. Biol.*, **9**, R929–R932.
68. van Dijk,M. and Bonvin,A.M.J.J. (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.*, **37**, W235–W239.

Extremophiles (2013) 17:881–895
DOI 10.1007/s00792-013-0586-9

REVIEW

Host and viral transcriptional regulators in *Sulfolobus*: an overview

Patrizia Contursi · Salvatore Fusco · Danila Limauro · Gabriella Fiorentino

Received: 30 July 2013 / Accepted: 16 September 2013 / Published online: 2 October 2013

Abstract The genus *Sulfolobus* includes microorganisms belonging to the domain Archaea, sub-kingdom Crenarchaeota, living in geographically distant acidic hot springs. Their adaptation to such particular habitats requires finely regulated mechanisms of gene expression, among which, those modulated by sequence-specific transcription factors (TFs) play a key role. In this review, we summarize the current knowledge on the repertoires of TFs found in *Sulfolobus* spp. and their viruses, focusing on the description of their DNA-binding domains and their structure–function relationship.

Keywords Transcriptional regulators · *Sulfolobus* spp. · *Sulfolobus* viruses · Genomics

Introduction

The transcription machinery in Archaea has a chimeric nature with eukaryote-like basal factors and bacterium-like regulative components (Bell et al. 2001). Transcription initiation is mediated by two general factors, TATA element binding protein (TBP) and transcription factor B (TFB) that bind the promoter region at the TATA box, and the B recognition element (BRE), respectively, determining the directionality of the transcription (Bell et al. 1999). A single RNA polymerase (RNAP) of 13 subunits (Wojtas and Abrescia 2013) is recruited to the TBP/TFB pre-initiation complex and initiates transcription at the transcription start site (TSS).

A recent in silico study from the group of Perez-Rueda analyzed the distribution of TFs identified in 52 archaeal genomes and showed that 53 % of TFs had at least one homolog in bacterial genomes, another 45 % was archaeal specific, less than 2 % had homology with only eukaryotes and 6 % was common in the three domains. The same study also showed that archaeal TFs are numerically underrepresented if compared to those of bacterial genomes with a similar genome complexity and that they are mainly constituted only by the DNA-binding module without additional regulatory domains. Such difference might be compensated by the employment of not yet known strategies of transcriptional regulation (Pérez-Rueda and Janga 2010).

The chimeric nature of the transcriptional apparatus raises interesting questions regarding the molecular mechanisms of transcriptional regulation that is how bacterial-like regulators may operate in an eukaryal-like basal context. In this regard, structural and functional characterization of TFs along with their identification in sequenced genomes represents a remarkable approach to expand the knowledge of the archaeal transcriptional regulation.

The action mode of archaeal transcriptional repressors is based on simple models of steric hindrance of TATA box, BRE or TSS (Bell 2005). Conversely, operator sequences located upstream of the TATA box or BRE can facilitate the recruitment of TBP and/or TFB to promoters (Bell and Jackson 2000). Furthermore, alteration of DNA conformation, like bending of the promoter region (Peeters et al. 2013; Vassart et al. 2013), or the use of alternative pairs of basal transcription factors (TBP–TFB) is another possible mechanism of transcriptional regulation (Tenorio-Salgado

Communicated by S. Albers.

P. Contursi · S. Fusco · D. Limauro · G. Fiorentino (✉)
Dipartimento di Biologia, Università di Napoli Federico II, via Cinthia, Edificio 7, 80126 Napoli, Italy
e-mail: fiogabri@unina.it

et al. 2011). The activity of TFs can be further modulated by chemical-physical agents, such as binding to small molecules, temperature shifts, pH variations, UV exposure (Bartolucci et al. 2013, Pedone et al. 2004).

The genus *Sulfolobus* belongs to the sub-kingdom Crenarchaeota, living in acidic hot springs around the globe with optimal growth temperatures of ~80 °C and pH's of ~3 (Brock et al. 1972). To date, the complete genome sequencing of several *Sulfolobus* species has been performed (Jaubert et al. 2013; Guo et al. 2011; Reno et al. 2009; Chen et al. 2005; She et al. 2001). Their comparative genomic analysis is functional to understand the evolution of the gene regulation networks that developed in distant habitats and to assess the biogeographical distribution of genome variation.

A survey of the extrachromosomal elements in *Sulfolobus* has revealed that this genus is a suitable host for viruses and cryptic as well as conjugative plasmids (Prangishvili et al. 2001; Lipps 2006; Prato et al. 2006, 2008). Members belonging to the *Fuselloviridae*, *Betalipothrixviridae*, *Rudiviridae*, *Guttaviridae* families of archaeal viruses infect different *Sulfolobus* species with the Fuselloviridae being the most representative (Peng et al. 2012; Pina et al. 2011). Indeed, they are also called *Sulfolobus* spindle-shaped virus (SSV). Transcriptional studies on SSV1, the best-studied member of this family, have an emblematic importance because they represent a milestone for the definition of the main components of the transcriptional machinery in the Archaea (Reiter et al. 1987, 1988).

Here, we provide an overview of the findings derived from studies of gene regulation in *Sulfolobus* and its viruses and plasmids from the perspective of the TFs. The main focus is the description of the DNA-binding domains adopted by the TFs in *Sulfolobus* as derived from the analysis of the databanks. A survey of the TFs structurally and/or functionally characterized, among which some investigated by our group will be also provided.

The DNA-binding domains in *Sulfolobus* TFs

An in silico inspection of TFs annotated in *Sulfolobus* genomes and its viruses showed that 4 main folds are employed to bind the DNA: the Helix–Turn–Helix (HTH), the Ribbon–Helix–Helix (RHH), the swapped hairpin and the C2H2 Zn finger.

The HTH domain is the most abundant DNA-binding structure described so far in prokaryotes and is considered to be originated from a common ancestor (Sauer et al. 1982; Aravind et al. 2005). The simplest version of the HTH domain is constituted by a basic HTH motif and a third helix at the N-terminus. The 2nd and the 3rd helices constitute the HTH motif, the critical determinant for DNA interaction (Fig. 1). The turn is sharp and the 3rd helix, known as the recognition helix, typically inserts into the major groove of the DNA, thus representing the main DNA–protein interface (Brennan and Matthews 1989). Commonly, the 3 helices of the domain are arranged in an approximately triangular shape. Generally, HTH proteins bind as dimers to palindromic DNA sequences in which each monomer recognizes a half site. Elaborations of the basic domain consist in extensions at the N-terminus or C-terminus. Examples are the tetra-helical version, characterized by an additional helix at the C-terminal found in AraC and TetR family members (Ramos et al. 2005), the multi-helical version with two additional helices at the N-terminus of the tetra-helical domain (TFIIB) or the winged HTH domain (wHTH) containing a wing at the C-terminus made, in the simplest version, of two antiparallel β-strands (β2 and β3) (Fig. 1) (Gajiwala and Burley 2000).

The superfamily of winged helix DNA-binding proteins includes the majority of *Sulfolobus* TFs. In these proteins, the wing often provides an additional interface for DNA contact, typically by interacting with the minor groove of DNA through charged residues in the hairpin (Gajiwala and Burley 2000).

Besides the simplest two-strand version, additional secondary elements of the wHTH domain can include β-sheets with 3-strands or 4-strands. The two- and three-stranded versions of the wHTH are present in some of the largest families of transcription factors (Lrp/AsnC and ArsR) as well as several eukaryotic DNA-binding domains. The Ferric uptake regulator (Fur) and Multiple-antibiotic resistance regulator (MarR) families of TFs involved in metal and drug transcriptional regulation, respectively, are characterized by a 2 or 3 stranded β-sheet with an additional helix at the C- terminus (Fig. 1f) (Perera and Grove 2010). A derivative of the 3 stranded version of the wHTH has been found in the DNA-binding domain of MerR transcriptional regulators. These proteins have lost the N-terminal helix which has been substituted by the first β-strand, but no structure is currently available from *Sulfolobus* spp. (Schelert et al. 2006).

To date, sequence comparisons and phylogenetic analyses focused on the HTH DNA-binding domain allowed the identification of different TF families through the determination of specific signatures for each family, whereas analyses of the regions outside the DNA-binding domain helped in the definition of the subfamilies (Rigali et al. 2002). The HTH domains can be combined in the same protein with other domains (for instance the RAM and TRASH domains) or organized in oligomeric assemblies giving rise to an amazing variety of architectures.

The RHH domain is characterized by the following order of the secondary structure elements: β-strand–α-

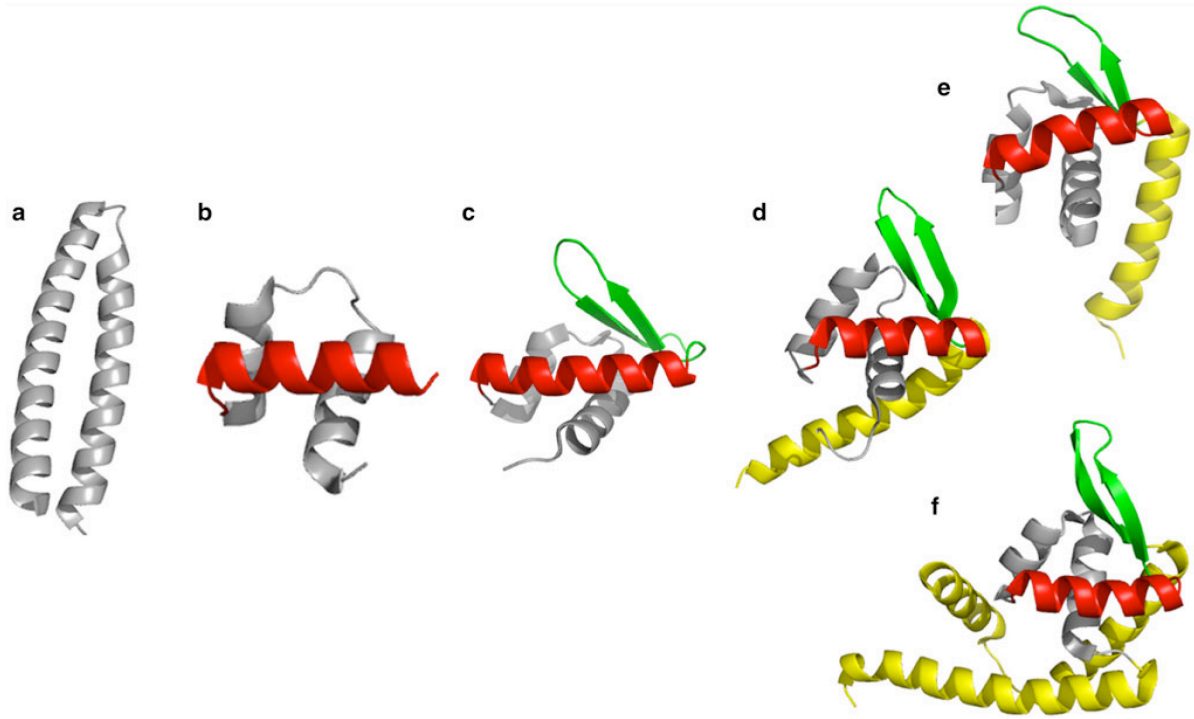


Fig. 1 HTH domain and its elaborations found in structures of TFs from *Sulfolobales*. The proteins are **a** D63 from SSV1 (Kraft et al. 2004a), **b** Lrp from *S. tokodaii* (Kumarevel et al. 2008a, b; Nakano et al. 2007), **c** F112 from SSV1 (Menon et al. 2008), **d** F93 from SSV1(Kraft et al. 2004b), **e** F-93 from STIV (Larson et al. 2007a), **f** BldR from *S. solfataricus* (Di Fiore et al. 2009; Fiorentino et al. 2007) (see also Tables 1, 2). *Gray* helix-turn-helix, *Red* recognition helix, *Green* strands making the wing. *Yellow* additional helix at the C-terminus

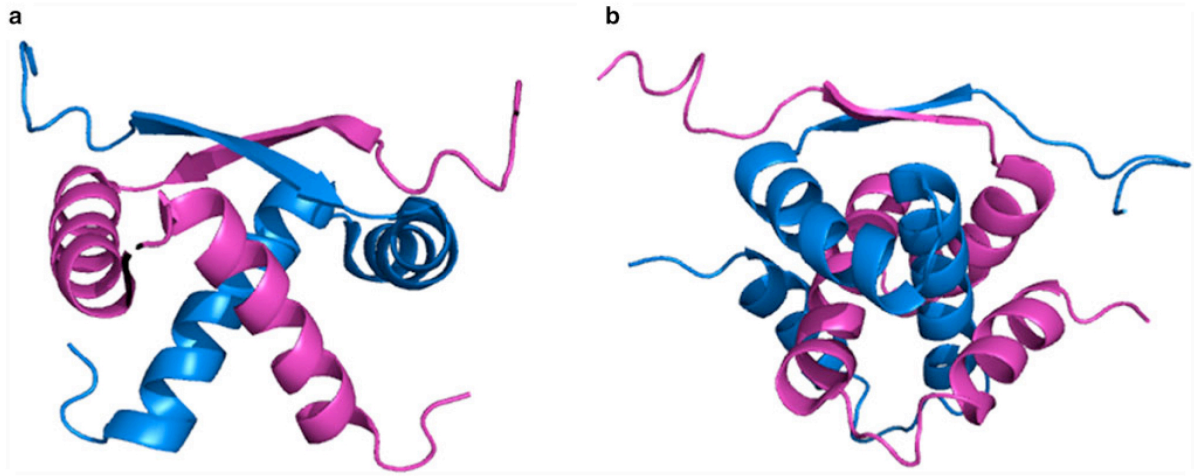


Fig. 2 RHH domains found in structures of TFs from *Sulfolobales*. The proteins are **a** ORF56 from pRN1 plasmid (Weininger et al. 2009) and **b** E73 from SSV RH (Schlenker et al. 2012) (see also Table 2). Each monomer is colored in *magenta* or *blue*

helix–α-helix; according to Aravind et al., it may be considered a distinct version of the HTH DNA-binding domain, with the main difference between the HTH and RHH transcription factors being in the topology of the DNA-binding determinants (Aravind et al. 2005); in the functional quaternary structure, two RHH motifs form compact and symmetric dimers $(RHH)_2$ exhibiting a two-fold axis centred on the two-stranded anti-parallel β-sheet

(Fig. 2). Nucleotide bases located on the major groove of DNA are contacted by residues of this sheet to form sequence-specific hydrogen bonds, whereas the first helix of the motif is involved in non-specific stabilizing phosphate backbone interactions. On the other hand, the C-terminal helix is mainly engaged in the inter-monomeric contacts of the quaternary $(RHH)_2$ structure (Schreiter and Drennan 2007).

Because of the small size of the domain and the lack of conservation of amino acid positions among RHH members, the identification via bioinformatics of new RHH sequences might be not always straightforward. However, a conservation pattern correlating with both DNA interactions and structure-stabilizing residues has been identified (Schreiter and Drennan 2007). Based on these structural features, a number of RHH proteins have been identified in all Archaea (including *Sulfolobus* spp.) using sensitive iterative research programmes such as position-specific iterated blast (PSI_BLAST) and Conserved Domains Database (CDD) (Prangishvili et al. 2006).

The RHH motif characterizes a functionally diverse superfamily of TFs, mainly repressors that down-regulate the expression of their own genes as well as of genes that are involved in many distinct biological processes, such as the uptake of metals (NikR), amino acid biosynthesis (MetJ, PutA), the control of plasmid copy number (CopG, Omega), the lytic cycle of bacteriophages/viruses (Arc, Mnt) and plasmid partitioning and conjugation (ParG/B, TrwA, TraY) (Schreiter and Drennan 2007; Guillière et al. 2009). This wide variety of functions is in part related to the fact that the RHH motif can be present in a number of different structural versions, i.e., either as an isolated RHH motif (as in Arc and CopG) or as a RHH repeat (TraY) or as part of larger proteins with additional domains (NikR). The genes encoding putative DNA-binding proteins bearing the RHH motif are well represented in *Sulfolobus* genomes and its plasmids as well as in the genomes of the viruses that infect them, thus pointing out a relevant role of these proteins in the regulation of transcription for both host and viral genes (Guillière et al. 2009). Interestingly, whereas all the crenarchaeal viral genomes contain at least one RHH-domain protein, the euryarchaeal ones encode only few proteins with a RHH domain. Such difference might be related to the highly compact and thermostable structure that makes the RHH proteins particularly suitable for functioning at high temperature (Peixeiro et al. 2013).

The third DNA-binding domain well represented in *Sulfolobus* genomes is the swapped hairpin fold (Prangishvili et al. 2006) found in homodimeric TFs: each monomer consists of a pair of β hairpin linked by a α-helix in the following sequence of the secondary structure elements: β1–β2–α–β3–β4. In this fold, the GD box, a Gly-Asp rich motif that allows an orthogonal turn of the α helix

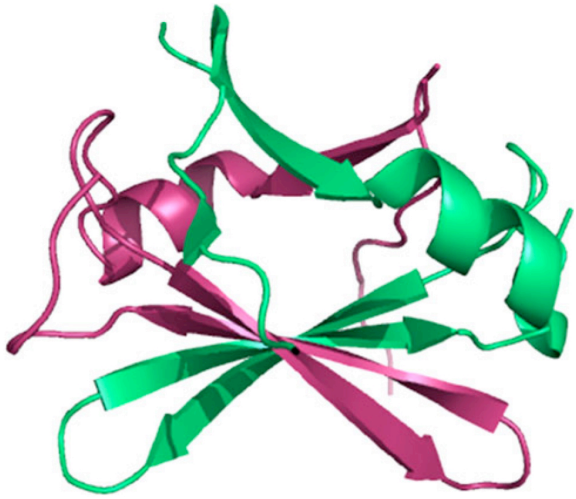


Fig. 3 Swapped hairpin fold found in the structure of protein c68 from the plasmid pSSVx (Contursi et al. 2011). Each monomer is colored in *magenta* or *green*

with respect to the contiguous β strands, connects the C-terminal end of the α-helix with the N-terminal of β3 Fig. 3) (Coles et al. 2005; Huffman and Brennan 2002). In the functional quaternary structure, the four β-hairpin elements are engaged in extensive hydrophobic interactions at the dimer interface; for this reason, the overall eight-stranded architecture is called swapped hairpin β-barrel domain (Liu et al. 2005). The swapped hairpin fold and its structural variants feature a large superfamily of prokaryotic TFs that include the key transition state regulator of *Bacillus subtilis* AbrB, SpoVT, PrlF from proteobacteria, MazE which functions as an antidote for the chromosomal addiction module MazE/MazF and plays a role in bacterial cell death, the plasmid maintenance proteins (VagC and VapB from *Actinobacillus actinomycetemcomitans and Salmonella,* respectively), and the archaeal PhoU homologs (Coles et al. 2005). This latter family includes metalloproteins involved in phosphate-specific transport; they are constituted by an N-terminal swapped hairpin fold and two repeats of a three helix bundle at the C-terminus involved in metal binding (Liu et al. 2005). Besides the PhoU homologs, all the other members of this superfamily are multi-domain proteins that contain a swapped hairpin N-terminal domain, and a C-terminal domain responsible for the oligomerization, the binding to partner proteins and/or small molecules (Contursi et al. 2011).

Searching for AbrB-like protein in *Sulfolobus* genomes retrieved about 200 sequences clustered into three length group, i.e., (1) those about 50–70 aa long, (2) those about 80–100 aa long and (3) those extending up to about 300 aa. The first group presumably includes proteins constituted by the only DNA-binding domain, the second one comprises

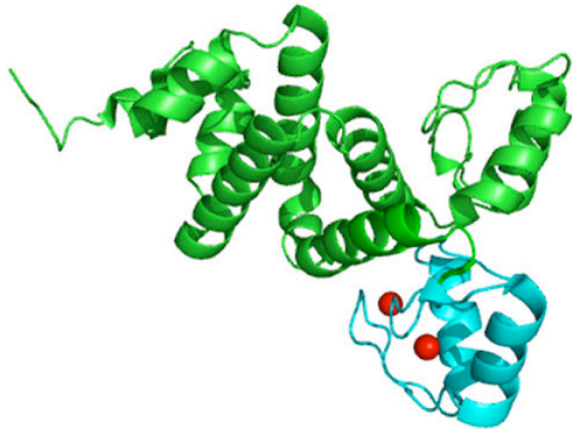


Fig. 4 B129 from SSV1 displays the C2H2 Zinc finger fold (PDB code 2WBT). Zinc ions are colored in *red*

two-domain proteins (similar to AbrB, SpoVT, etc.), and the third one is mainly represented by the PhoU homologs.

Zn fingers-like proteins represent an intriguing feature of the *Sulfolobus* TFs collection, because this class of proteins has been found mainly in eukaryotic transcriptional proteins and only in putative TFs from prokaryotic plant pathogens (Prangishvili et al. 2006, Malgieri et al. 2007).

The Zn finger proteins, in which the zinc plays mainly a structural role, perform a broad range of functions such as replication, transcription, translation, metabolism, etc. Zn finger structures fall into eight folds, but those found in transcriptional regulators are: the C2H2-like finger, the zinc ribbon and the Zn2/Cys6 finger folds (Krishna et al. 2003; Chen et al. 2000).

Searching for Zn finger proteins in *Sulfolobus* genomes revealed a remarkable abundance of proteins bearing only C2H2-like finger and Zn-ribbon folds. The C2H2 type motif has also been found in viral putative transcription factors, such as B129 from SSV1 and AFV1p06 from *Acidianus* filamentous virus 1 (Guillière et al. 2013). It is constituted by a β-hairpin followed by an α-helix that forms a left-handed ββα-unit; of the two zinc ions, one is coordinated by a zinc knuckle (a unique turn with the consensus sequence CPXC) located at the end of the β-hairpin and the other is bound from two ligands coming from the C-terminal of the α-helix (Krishna et al. 2003) (Fig. 4). In the Zn ribbon fold, two zinc-knuckles are involved in binding of the two Zn ions. The core of the structure is composed by two β-hairpins forming two structurally similar zinc binding sub-sites. Although Zn-ribbon proteins are well represented in *Sulfolobus* genomes, no TFs have been structurally characterized so far. In most Archaea, the Zn-ribbon is found fused to HTH domains (Aravind and Koonin 1999) and the proteins bearing such associated domains are predicted to be transcription regulators unique to Archaea. However, this association is missing in *Sulfolobus* species.

TFs in *Sulfolobus* spp.

Lrp family members

The Lrp family, referred also as AsnC or Feast Famine Regulatory Proteins (FFRP), comprises TFs that are widely distributed among prokaryotes and are characterized by a highly conserved structure (Brinkman et al. 2003; Calvo and Matthews 1994). They are typically 15 kDa monomeric proteins, with a wHTH domain located at their N-terminus linked by a β strand to the Regulation of Amino Acid Metabolism (RAM) domain at C-terminus. The RAM domain consists of a characteristic αβ sandwich fold involved in the recognition of the effectors (Ettema et al. 2002) and possibly in oligomerization. Lrp regulators bind in proximity of the promoter regions of their own and target genes and act as activators or repressors depending on the specific case.

Although the role of most Lrp-like regulators in *Sulfolobus* must still be elucidated, their high occurrence in *Sulfolobus* genomes suggests an important role to coordinate the metabolic network and guarantee the optimal performance under stress conditions (Peeters and Charlier 2010). Members characterized from *Sulfolobus* spp. bind an array of ligands and are involved in amino acid metabolism, in transport and/or in cell energy processes, highlighting their flexibility and global role. In case they are activators, the stimulation of transcription takes place or by binding upstream the BRE and TATA boxes or by DNA bending and/or wrapping induced by Lrp oligomerization (Peeters et al. 2004). Conversely, repression is mediated by Lrp interaction with a specific ligand and/or by interference with the formation of the pre-initiation complex. To date, eight Lrp-like TFs have been characterized in *Sulfolobus* spp. and their main features are reported in Table 1.

The first Lrp-like protein, Lrs14, was discovered in *S. solfataricus* and it is an autorepressor distantly related to the other members of the Lrp family. Neither leucine nor asparagine affects Lrs14 binding to its own promoter; therefore, a possible role as a global regulator involved in the adaptation to different nutritional conditions has been suggested (Napoli et al. 1999; Bell and Jackson 2000).

SsLrp shows 30 % identity with Lrp of *E.coli* and represents the archaeal counterpart to bacterial Lrp (Enoru-Eta et al. 2000; Enoru-Eta et al. 2002). It autoregulates its own transcription preventing promoter recognition by TBP and TFB (Enoru-Eta et al. 2002).

Table 1 TFs experimentally described in *Sulfolobus* spp.

Family	*Sulfolobus* spp.	Member	Genome annotation	UniProt number	PDB number	Oligomeric state	Function	References
Lrp/AsnC	*S. solfataricus P2*	Lrs14	*Sso1108*	O93774	–	Dimer	Repressor	Napoli et al. (1999), Bell and Jackson (2000)
		SsLrp	*Sso0606*	Q7LXS6	–	Tetramer	Putative Repressor	Enoru-Eta et al. (2002)
		SsLrpB	*Sso2131*	Q7LX67	–	Dimer	Activator	Peeters et al. (2004, 2009, 2013)
		LysM	*Sso0157*	Q980W9	–	n.d.	Putative activator	Brinkman et al. (2002), Song et al. (2013)
	S. solfataricus 98/2	CopT	*Sso2652*	Q97VH3	–	n.d.	Putative repressor Activator	Ettema et al. (2006), Villafane et al. (2009)
		CopR	*Ssol_0462*	B2CQV6	–	n.d.		
	S. acidocaldarius	Sa-Lrp	*Saci1588*	F2Z5Y7	–	Tetramer/ Octamer	Putative activator/ repressor	Enoru-Eta et al. (2000), Vassart et al. (2013)
	S. tokodaii	Grp	*St1022*	F9VNT4	2E7X	Octamer	–	Kumarevel et al. (2008a), Nakano et al. (2007)
MarR	*S. solfataricus P2*	BldR	*Sso1352*	Q97YG9	3F3X	Dimer	Putative activator	Di Fiore et al. (2009), Fiorentino et al. (2007)
		BldR2	Sso1082	Q97Z53		Dimer	Putative repressor	Fiorentino et al. (2011)
	S. tokodaii	St1710	*St1710*	Q96ZY1	3GF2	Dimer	–	Kumarevel et al. (2008b, 2009)
ArsR	*S. solfataricus P2*	SsMerR	*Sso2688*	Q97VE0	–	Dimer	Repressor	Schelert et al. (2004, 2006)
FHA	*S. tokodaii*	ST0829	*St0829*	F9VN98	–	n.d.		Duan and He (2011)
Unclassified	*S. solfataricus P2*	Sta1	*Sso0048*	Q981A1	NMR	Dimer	Activator	Kessler et al. (2006)
SpoVT/ AbrB	*S. solfataricus P2*	SSo7c4	*SSo0046.1*	P81552	NMR	Dimer	Putative repressor	Hsu and Wang (2011)

Ss-LrpB and LysM have been described as the most flexible regulators. The former cooperatively binds to three distinct semipalindromic sequences included in its own promoter (Peeters et al. 2009). Other targets described for this protein are the promoters of the pyruvate ferredoxin oxidoreductase (POR) operon (*Sso2128-Sso2129-Sso2130*) and of two permease genes (*Sso2126* and *Sso2127*) located in its genomic environment; upon binding of Ss-LrpB, the expression of all of these genes is activated. From a functional point of view, Ss-LrpB is distinct from the other Lrp-like proteins, because it could have a key role in cell energy processes more than in the amino acid metabolism regulation; this latter role was excluded by the lack of interaction with any amino acid (Peeters et al. 2009; Yokoyama et al. 2009).

LysM is involved in the regulation of lysine biosynthesis. The biosynthetic genes are organized in two consecutive operons *lysYZM* and *lysWXJK* showing the same polarity. The operons are transcribed from two promoters, one upstream *lysY* and the other upstream *lysW*. LysM binds upstream the BRE and TATA box of *lysW* promoter stimulating transcription and in the presence of lysine the binding efficiency is reduced (Brinkman et al. 2002; Song et al. 2013). Recently, other cofactors (L-arginine, L-homoarginine, L-glutamine, L-methionine, L-leucine, L-isoleucine and L-valine) have been found affecting the LysM binding capability not only to target promoters, but also to the coding regions of different genes involved in amino acid biosynthesis and transport. This last aspect suggests a possible control of LysM concentration inside the cell (Song et al. 2013).

Other members of the Lrp family are Sa-Lrp and the Glutamine receptor protein (Grp), characterized in *S. acidocaldarius* and *S. tokodaii*, respectively.

Sa-Lrp binds the glutamine as effector which induces the formation of an octamer with enhanced DNA-binding specificity to its own promoter. As reported by Vassart et al., Sa-Lrp controls also the transcription of genes involved in nitrogen metabolism and in UV-induced pili synthesis (Vassart et al. 2013).

Grp from *S. tokodaii* has been solved both in the ligand free and in the ligand-bound forms. Glutamine, its effector, determines the formation of the octamer which facilitates gene activation through DNA bending and wrapping (Kumarevel et al. 2008a).

A copper regulator, CopT, involved in the control of copper homeostasis, has been functionally characterized in *S. solfataricus* P2. The *cop* operon involves copper sensing and response; *cop* loci encoding CopT, the transcriptional regulator, CopM a metallochaperone and CopA, a P-type copper-exporting ATPase have been found in *S. solfataricus, tokodai* and *acidocaldarius* (Bartolucci et al. 2013; Ettema et al. 2006). CopT has been assigned to the Lrp family because of the presence of an N-terminal wHTH domain. An additional putative metal-binding motif (TRASH domain) is located at the C-terminus and is constituted by a conserved cysteine involved in copper sensing (Ettema et al. 2003). CopT acts as a repressor of *copMA* operon by binding to multiple sites both upstream and downstream the BRE and TATA boxes and transcription takes place by copper-CopT derepression. Copper resistance and its regulation were analyzed also in *S. solfataricus* 98/2 strain (Villafane et al. 2009).

MarR family members

MarR family proteins constitute a diverse group of TFs that modulate the expression of genes encoding proteins involved in a wide variety of cellular processes including metabolic pathways, stress responses, virulence and degradation or export of harmful chemicals such as phenolic compounds, antibiotics and common household detergents (Perera and Grove 2010). All MarR proteins are able to bind either the DNA through a conserved wHTH motif (Fig. 1f; Table 1) or specific ligands (many not yet identified) which, upon binding, modulate DNA recognition.

In the genus *Sulfolobus*, members of this family might be critical for their adaptation to particular habitats or lifestyles. The exploration of homologous protein sequences (sharing sequence identity around 30 %) among *S. solfataricus, S. islandicus, S. tokodaii,* and *S. acidocaldarius*, revealed that, differently from the majority of bacteria, they all have at least two conserved MarR representatives, indicating the presence of several regulatory systems involved in multiple-antibiotic resistance. Many of these mechanisms are still uncharacterized, but in the last years some *Sulfolobus* MarR proteins have been deeply studied (Fiorentino et al. 2007, 2011; Di Fiore et al. 2009; Kumarevel et al. 2009).

In particular, our group recently characterized two members in *S. solfataricus*, the proteins BldR and BldR2, giving insights into their possible physiological regulative roles and the protein/ligand/DNA-binding mechanism. It was shown that BldR acts as an activator by binding to its own promoter and thus increasing the expression of a *MarR-like* operon and of an alcohol dehydrogenase gene (*adh*) (*Sso2536*) (Fiorentino et al. 2007, 2003; Contursi et al. 2003; Cannio et al. 1999). The accumulation of the ADH enzyme determines the conversion of the aldehydes to the less toxic corresponding alcohols (Fiorentino et al. 2007; Cannio et al. 1996). It was also demonstrated that the interaction with the effector benzaldehyde stimulated BldR binding to the target promoters; hence, similarly to bacterial counterparts, it responded to the effector molecule by direct binding. An in vivo study also showed that the aromatic aldehyde was the environmental signal/ligand for BldR; in fact upon expression of BldR and benzaldehyde addition, an increase in heterologous *gfp* gene expression driven by the *adh* promoter was revealed (Fiorentino et al. 2009). The crystallographic structure of BldR at 1.90 Å resolution was solved, and used as a starting point for a molecular modeling study on the BldR-DNA adduct. Molecular dynamics supported previous hypotheses that the ability of MarR family members to act as activators or repressors could be not related to a particular DNA-binding mechanism but is essentially due to the position of the specific binding site on the target DNA (Di Fiore et al. 2009).

BldR2 is a highly thermostable dimeric protein which binds specifically to its own promoter at two palindromic sites overlapping BRE-TATA and the transcription start site. According to the location of the identified basal promoter elements and the BldR2 binding region, it was proposed, differently from BldR, that the protein could act as a repressor. Benzaldehyde and salicylate in fact were able to weaken the interaction of BldR2 with its target, suggesting that salicylate and to a lesser extent benzaldehyde could be ligands for BldR2; but the low ligand affinity measured raised questions about their physiological relevance and suggested that other aromatic compounds, possibly derived from aromatic catabolic pathways, could be the natural effectors. Hence, BldR2 in vivo would control regulatory mechanisms diverse from those regulated by BldR (Fiorentino et al. 2011).

The crystal structure of a close relative of BldR, ST1710 from *S. tokodaii* was also determined in three different forms: apo-form, ST1710-salicylate and ST1710-DNA complex (Table 1). The ST1710-DNA complex shares the topology of apo-ST1710 (Miyazono et al. 2007; Kumarevel et al. 2008b) and of other MarR proteins. The study outlined that whereas the ligand binding did not affect the overall structure, a significant conformational change occurred upon binding to the DNA (Kumarevel et al. 2009).

Other *Sulfolobus* TFs

A few other TFs, with distinct structural or functional features and/or less common DNA-binding folds, have been described.

SsMerR from *S. solfataricus* is a member of the MerR family characterized by a wHTH domain located at the N-terminus linked by a coiled-coil region to the C-terminal

part bearing the Hg(II) binding site. SsMerR from *S. solfataricus* is phylogenetically distinct from the majority of the bacterial counterparts because of its shorter length and the presence of only two cysteine residues generally required for ligand-induced DNA release, instead of three (Schelert et al. 2004). SsMerR acts as a repressor of *merHAI* operon by binding to an inverted repeat that overlaps the BRE sequence in the *merH* promoter. Hg^{2+} derepresses the transcription interacting with SsMerR (Schelert et al. 2006).

Sulfolobus transcription activator1 (Sta1) is a regulatory protein encoded by the chromosome of *S. solfataricus* and acts as an activator through the binding at viral promoters of *S. islandicus* rod-shaped virus 1 (SIRV1). The NMR structure indicated that the protein contains a wHTH DNA-binding domain consisting of a two- or three-stranded antiparallel β-sheet and three α-helices. It has been proposed that the recruitment of a cellular regulator of different nature by the virus could represent a viral tactic to promote viral gene transcription (Kessler et al. 2006).

Another putative TF, whose structure has been solved by NMR, is Sso7c4. This protein bears a homodimeric DNA recognition fold, typical of the SpoVT-AbrB superfamily and called swapped β-loop-β "Thai Chi" for the symmetric arrangement of the two β1-loop-β2 and β′1-loop-β′2 motifs (Hsu and Wang 2011). Sso7c4 binds to double-stranded DNA, but a specific target has not been identified. Based on sequence homologies with two transcriptional repressors of *E. coli*, the repressor protein ACCR in *Agrobacterium tumefaciens* and the initiation factor aIF-1A from *Methanococcus jannaschii*, a role as transcriptional repressor has been hypothesized (Oppermann et al. 1998).

Recently, another uncommon TF has been characterized from *S. tokodaii*: ST0829, ascribed to the FHA transcription factor family. It contains an N-terminal Zn finger-RanBP domain, which is responsible for DNA binding and a phosphopeptide recognition domain, the FHA domain, at the C-terminus. ST0829 was found to interact with the promoter (ST2519p) of the operon *flaGFHIJ* which encodes flagellar proteins; its binding is negatively regulated by phosphorylation of Ser/Thr protein kinase ST1565 (Wang et al. 2010). The transcriptional regulation of flagellar proteins in *S. tokodaii* suggests that phosphorylation could play an important role in the environmental adaptation (Aivaliotis et al. 2009); in fact archaeal flagella are responsible not only for motility but also for cell–cell interaction and surface adhesion that can assure the survival in harsh habitats (Duan and He 2011).

Sulfolobus viral and plasmidic TFs

Viruses infecting *Sulfolobus* spp. show unique capsid morphology as well as peculiar genome sequences and organization. Indeed, their classification has required the introduction of new viral families, i.e., *Fuselloviridae*, *Guttaviridae*, *Lipothrixviridae* and *Rudiviridae* (Prangishvili et al. 2001), although some still remain unclassified (Pina et al. 2011). For several of them the entire genome has been sequenced and comparison analyses showed little similarity to DNA sequences deposited on public databases (Palm et al. 1991; Stedman et al. 2003; Rice et al. 2001, 2004; Peng et al. 2001; Arnold et al. 2000; Prangishvili et al. 1999; Xiang et al. 2005, Redder et al. 2009); hence, functional assignments of the predicted gene products have been considerably hindered (Wiedenheft et al. 2004; Prangishvili et al. 2006; Stedman et al. 2003).

With the purpose of elucidating their role in the viral life cycle and to identify new folds, a recent high-throughput approach has been performed by the Scottish structural proteomics facility on several *Sulfolobus islandicus* rudivirus (SIRV1) proteins. The crystal structures are available at the Protein Data Bank (PDB) website (http://www.rcsb.org/pdb/home/home.do) with the following codes: ORF131 (2X5G/H), ORF55 (2X48), ORF119 (2X3G) and ORF114 (2X4I). Although these analyses led to the successful determination of protein structures, the lack of functional information still hampers the assignment of a clear biological role (Oke et al. 2010).

Conversely, structural and functional features are available for several gene products encoded by the genomes of the fusellovirus SSV1 and for some proteins of STIV. They are summarized in Table 2.

TFs from *Sulfolobus* spindle-shaped virus 1

SSV1 is an UV-inducible fusellovirus isolated from *Sulfolobus shibatae* cells in Beppu (Japan) (Martin et al. 1984) and its genome has been the first to be sequenced (Palm et al. 1991).

Among the 36 predicted SSV1-encoded proteins, structural and/or functional data are available for only four TFs, i.e., B129 (2WBT) (Lawrence et al. 2009), D63 (1SKV) (Kraft et al. 2004a), F93 (1TBX) (Kraft et al. 2004b) and F112 (2VQC) (Menon et al. 2008). While for B129 and D63, only the structural features are available, a functional characterization of F93 and F112 has been reported.

B129 contains the C2H2 zinc-finger motif (Fig. 4), while D63 is a HTH protein (Fig. 1a) that forms a homodimeric four-helix bundle stabilized by an extensive hydrophobic interface, intermolecular salt bridges and hydrogen bonds. Although the structure of D63 has been solved at atomic level, a reliable assignment of its function was precluded since both the HTH motif and the four-helix bundle topology can be also adopted by many proteins with

Table 2 TFs experimentally described in plasmids and viruses of *Sulfolobus* spp.

Virus or Plasmid	Protein	Fold	NCBI number	UniProt number	PDB number	Oligomeric state	Function	Reference
SSV1	D63	HTH	NP_039786.1	P20215	1SKV	Homodimer	Putative TF	Kraft et al. (2004a)
SSV1	F93	wHTH	NP_039783.1	P20222	1TBX	Homodimer	Putative TF	Kraft et al. (2004b)
SSV1	F112	wHTH	NP_039787.1	P20220	2VQC	Monomer	Putative TF	Menon et al. (2008)
SSV1	B129	Zn finger	NP_039795.1	P20201	2WBT	Homodimer	Putative TF	Not published
SSV1	F55	RHH	DAA64513.1	–	–	Homodimer	Putative NAR	Fusco et al. (2013)
SSV RH	E73	RH3	NP_963940.1	–	4AAI	Homodimer	Putative TF	Schlenker et al. (2012)
SIRV1	SvtR	RHH	2KEL_A	–	2KEL	Homodimer	Putative TF	Guillière et al. (2009)
STIV	F-93	wHTH	2CO5_A	–	2CO5	Homodimer	Putative TF	Larson et al. (2007a)
STIV	B116	Saddle shape	2J85_A	–	2J85	Homodimer	Putative TF	Larson et al. (2007b)
pRN1	ORF56	RHH	NP_044371	–	2K9I	Homodimer	Putative NAR	Weininger et al. (2009)
pSSVx	C68	Swapped hairpin	CAB81816.1	Q9P9J8	3O27	Homodimer	Putative PAR	Contursi et al. (2011)

TF transcription factor, *NAR* negative auto regulator, *PAR* positive auto regulator

a multitude of functions (Kraft et al. 2004a). Nevertheless, the structural similarity of D63 with the adaptor protein Repressor of Primer (ROP) of *Escherichia coli*, suggested a similar function in the copy number regulation of the SSV1 episomal DNA (Banner et al. 1987; Helmer-Citterich et al. 1988; Eberle et al. 1991).

F93A is a more complex HTH protein (Fig. 1d); its polypeptide chain shows three α-helices followed by two β-strands and finally ends with a fourth α-helix (α1–α2–α3–β1–β2–α4) (Kraft et al. 2004b). In solution, it assembles in homodimers, whose structure shows homology to SlyA (Wu et al. 2003) and MarR (Alekshun et al. 2001) subfamilies of wHTH DNA-binding proteins. Unlike the latters, F93 has a much simpler dimer interface which leads to rule out the involvement of this region in modulating its binding activity. Although electrostatic surface calculations allowed the identification of the putative DNA-binding surface, the capability of F93 to interact with specific SSV1 targets has not been demonstrated (Kraft et al. 2004b). F93 shares structural similarity with a protein of *Bacillus subtilis* that acts as terminator of the replication (Wilce et al. 2001), thus suggesting that F93 might have an alternative role in the replication of the SSV1 genome.

In a more recent work, it has been reported the structural and functional characterization of the SSV1 gene product F112 (Menon et al. 2008). The sequence of the secondary structure elements is: α1–β1–α2–α3–β2–β3, thus defining a further elaboration of the wHTH fold (Fig. 1c). Indeed, a short β-strand connects the two helices of the HTH motif. Interestingly, the monomeric structure of F112 is stabilized by an intramolecular disulfide bond, as often observed for thermophilic proteins. As for F93, the functional target of F112 has not been defined.

TFs from *Sulfolobus* turreted icosahedral virus (STIV)

The protein F-93 (PDB code 2CO5) of the virus STIV (Larson et al. 2007a) has structural similarity with the gene product F93 of the fusellovirus SSV1 (Kraft et al. 2004b); in fact, similarly to F93, it shows two variations with respect to a typical wHTH fold: (1) a small β-strand (β1) inserted between the helices α1 and α2, and (2) an additional C-teminal α-helix (α4) (Fig. 1e). In most cases, C-terminal extensions are involved in the dimerization and F-93 is not an exception to this role. In this case, a cysteine residue at the end of the C-terminal helix (Cys93) forms an intermolecular disulfide bond. Both the structure and the peculiar localization of the ORF encoding F-93 on the STIV genome led to speculate that this protein might act as a transcriptional regulator for other viral genes (Larson et al. 2007a).

In a recent structural analysis of the STIV proteome undertaken by Larson et al. (2007b), the crystal structure of the protein B116 has been reported. Although the fold of this protein is not related to any of the DNA-binding domain described above, the description of the structural features of B116 is included in this review, because a role in DNA-binding has been hypothesized. Indeed, each protein monomer consists of five-stranded ß-sheet flanked at one side by three α-helices (Fig. 5). A disulfide bond between Cys33 and Cys62 covalently connects the helices

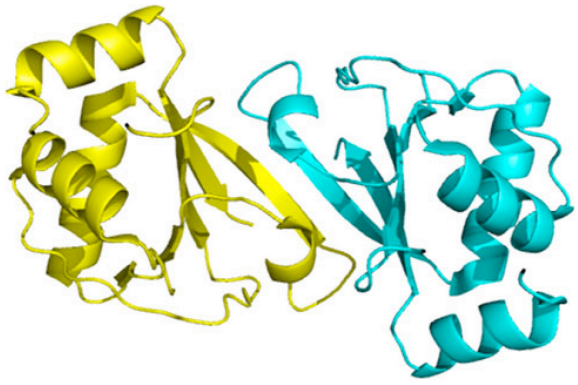


Fig. 5 B116 from STIV (Larson et al. 2007b) displays the unusual saddle shape fold

α1 and α2, thus providing further thermostability to the protein structure. Two monomers of B116 interact by means of their β-sheets in order to form a homodimer with a 10-stranded intermolecular β-sheet. The topology of this region, which represents the dimer interface, strongly resembles an unclosed β-barrel, while the remaining portion of the whole structure gives rise to a saddle-shaped protein. Interestingly, genomes of several crenarchaeal viruses, which belong to other three viral families, encode B116-like proteins. Since structural comparisons performed using the DALI and VAST were unsuccessful, only the analysis of conserved surface features served to speculate about the role of B116. Multiple alignments highlighted a strict conservation of positive-charged amino acids, which in the structure form a contiguous surface along the rim on its central cleft. This feature along with the absence of small pockets on the protein surface and its intrinsic twofold symmetry strongly indicate that B116 might interact with macromolecules that should also possess twofold symmetry and carry negative charges, such as the DNA, being possibly involved in transcription regulation or viral DNA packaging (Larson et al. 2007b).

Other viral/plasmidic TFs

In crenarchaeal plasmids and viruses, a considerable abundance of genes encoding for RHH proteins has been reported (Prangishvili et al. 2006).

One example is the protein SvtR of the *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1) that folds into a dimeric RHH; likewise, most of the RHH members also SvtR down-regulate its own expression. Another target of SvtR is the promoter of the *gp30* gene, which encodes a structural protein presumably required in a later infection stage (Guillière et al. 2009).

Another well-known RHH transcription regulator is ORF56 encoded by the plasmid pRN1 of *S. islandicus* REN1H1 (Weininger et al. 2009). The corresponding gene overlaps the *orf904* gene (*repA*), which encodes the main protein involved in the initiation of plasmid replication. Such genomic arrangement is typical for copG regulators, which affect negatively the transcription efficiency of the *Rep* gene (as well as of its own gene) by binding to regulative sequences in the common *copG/repA*, and hence controlling indirectly the plasmid copy number (Lipps et al. 2001a).

F55 of the UV-inducible SSV1 is another example of a RHH regulator indirectly involved in the regulation of the episomal copy number. The protein fold has been deduced from secondary structure predictions and the recognition of the molecular determinants of RHH proteins (Schreiter and Drennan 2007). This newly identified transcriptional regulator specifically binds to operator sequences located in the promoter of the early genes as well as in its own promoter. F55 is mainly expressed in the absence of the UV stimulus, thus suggesting a more global regulation role for the lysogenic state of SSV1. It has been hypothesized that the regulation mode of F55 implies oligomerization starting from the primary binding sites and subsequently extending to the less affine in a concentration-dependent manner (Fusco et al. 2013). Many archaeal viral RHH transcriptional regulators could also share this mechanism of oligomerization (Peeters et al. 2013) that has indeed been demonstrated for AvtR, *Acidianus* filamentous virus 6 (Peixeiro et al. 2013). The protein E73 of the fusellovirus SSV RH (Ragged Hills) represents an interesting variation on the theme of the RHH regulators. Intriguingly, it shows a C-terminal extension consisting of an additional α-helix, thus giving rise to the RH3 motif (Fig. 2b). As the majority of the RHH proteins, the native E73 has a homodimeric architecture stabilized by inter-monomeric interactions mediated by the N-terminal β-strands of each monomer. The third α-helix is involved in the stabilization of the quaternary structure by encircling the helix α2 of the nearby subunit. This structural feature allows the formation of a tightly intertwined dimer whose dissociation occurs only upon a significant rearrangement of the helices α3 (Schlenker et al. 2012). Proteins that share sequence identity with E73 are encoded by other SSV genomes like SSV2 79a, SSV4 73, SSV5 GP23 and SSV6 GP17 and with E51 of SSV1, thus suggesting that this kind of fold is suitable for transcription regulation in the *Fuselloviridae* family.

Besides regulators such as HTH, RHH and their variant (wHTH and RH3), it has been recently reported the structural characterization of the protein C68 encoded by the plasmid-virus hybrid pSSVx (Arnold et al. 1999). C68 assumes the swapped hairpin fold, which is associated with

the prokaryotic transcription factors family of AbrB from *Bacillus subtilis* (Fig. 3). Two monomers of C68 interact with each other by a swapped hairpin mechanism leading to the formation of a dimer that is stabilized by several intermolecular interactions. The functional dimer is able to bind two operator sequences located in the promoter region of its own gene (Contursi et al. 2011). The temporal expression pattern of *c68* throughout the life cycle of pSSVx and its up-regulation led to speculate that it might act as positive transcriptional regulator (Contursi et al. 2006, 2007, 2010).

The most conserved gene in crenarchaeal conjugative and cryptic plasmids is *plrA* (ORF76 on the pSSVx genome and ORF80 on the pRN1 plasmid) (Contursi et al. 2007, Keeling et al. 1996). Results from G. Lipps group showed that ORF80 is a DNA-binding protein that binds to two palindromic sequences ($TTAAN_7TTAA$) located in its own promoter and spaced about 60 bp from each other. The function of this protein in vivo is currently obscure. However, a role in the binding to the replication origin of the pRN1 plasmid has been hypothesized based on the ability of ORF80 to assemble into large nucleo-protein complexes encompassing both the binding sites. Sequence analysis of ORF80 suggested that it might be a leucin-zipper-like protein (Lipps et al. 2001b). Our group carried out an NMR analysis of the ORF76 from pSSVx to unravel the structure of this protein (manuscript in preparation).

Concluding remarks

In the genus *Sulfolobus*, the elucidation of transcription regulation and the dissection of the complex regulatory networks have not developed to the point of substantial synergy due to the still limited current availability of tools for genetics, physiology, biochemistry, as well as structural and mechanistic analysis (Geiduschek and Ouhammouch 2005).

Nevertheless, a number of genome-wide approaches for the identification of TFs and their DNA-binding sites have been applied (Song et al. 2013; Nguyen-Duc et al. 2013). These evidences, together with the functional and structural characterization of a number of TFs from *Sulfolobus* and its viruses and plasmids, contribute to understand the basal transcription machinery and its interplay with global as well as gene-specific regulators of this archaeon.

In this work, we have updated the present functional and structural knowledge on *Sulfolobus* TFs and have analyzed their distribution in the different species.

In line with the representativeness of TFs in Crenarchaea, the most abundant families of TFs in *Sulfolobus* genomescapes are Lrp/AsnC, MarR and ArsR which all possess a wHTH DNA-binding domain. It has been speculated that the abundance of this fold is a consequence of its performance under the extreme conditions in which *Sulfolobus* spp. thrive as well as of its functional versatility. Alternatively, *Sulfolobus* might have expanded certain small size families, to generate a variety of different combinatorial assemblies (Pérez-Rueda and Janga 2010).

On the other hand, viruses and plasmids of *Sulfolobus* have a set of TFs displaying a wider repertoire of DNA-binding domains; in fact they belong mainly to one of the four fold groups, i.e., the RHH, the HTH, the swapped hairpin and the C2H2 zinc finger or the Zn-ribbon.

In line with the evidence that a major portion of archaeal TFs do not possess additional regulative effectors binding domains in addition to the DNA-binding module, also TFs from *Sulfololobus* viruses and plasmids are in general small and compact proteins mainly constituted only by the DNA-binding motif. Hence, how these small TFs can interact and/or interfere with the basal components of transcription machinery is still an open question, but preliminary evidences are emerging.

It has been proposed that TFs monomers can form high-ordered combinations (dimers/tetramers/octamers) with differing regulatory functions, as it has been shown for SaLrp (Vassart et al. 2013), or more sophisticated and alternative hetero-multimeric complexes by interacting with other TFs (Pérez-Rueda and Janga 2010).

Members belonging to the MarR family can exert their regulative role by being malleable at the dimer interface, and versatile in interacting with effectors of different chemical nature. Furthermore, since residues responsible for specific protein–DNA contacts are generally in limited number, it is likely that single-aminoacid substitutions could determine changes in the specificity of the binding site and/or a change from positive to negative regulation (Di Fiore et al. 2009).

The broader viral and plasmidic TFs repertoire is evident in the structural/functional variants identified in the RHH proteins of *Sulfolobus* viruses and could be a direct consequence of the quick pace at which viruses and plasmid evolve through horizontal gene transfer (from cells to viruses and vice versa), non-orthologous gene displacement, deletion/duplication events.

For instance, STSV1 encodes for a group of paralogous RHH-domain proteins that do not show any significant homology to any RHH protein and their genes are arranged in tandem, suggesting that they derive from intra-genomic duplication. Furthermore, SIRV and SIFV encode for a variant of the RHH proteins which are related to the plasmid partitioning proteins ParG/B.

Similarly, the resolution of the c68 protein from the hybrid virus/plasmid pSSVx and its functional characterization highlighted the exploitation of structural variants to make these TFs suitable for exerting its regulative role

through site-specific binding, instead of the promiscuous recognition of diverse target promoters typical of the closely related Abr-B members (Contursi et al. 2011).

Interestingly, the genomescapes of crenarchaeal viruses are mainly barren, with a large portion of genes coding for proteins that are not conserved in any other cellular form of life. Therefore, it is possible that some novel TFs with a lineage-specific nature will be identified and novel regulation mechanisms of gene expression will be unraveled.

In conclusion, the identification in the genomes of TF repertoires as well as their functional/structural characterization represents a valuable approach not only as a starting point for high-throughput studies aimed at determining their binding sites and their true physiological role, but also to get insight into the evolution of TFs in the three domains of life.

Acknowledgments We acknowledge Prof. Simonetta Bartolucci for valuable scientific discussions throughout the preparation of the manuscript.

References

Aivaliotis M, Macek B, Gnad F et al (2009) Ser/Thr/Tyr protein phosphorylation in the archaeon *Halobacterium salinarum* a representative of the third domain of life. PLoS ONE 4:e4777

Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. Nat Struct Biol 8:710–714

Aravind E, Koonin V (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. Nucl Acids Res 27:4658–4670

Aravind L, Anantharaman V, Balaji S et al (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev 29:231–262

Arnold HP, She Q, Phan H, Stedman K, Prangishvili D, Holz I, Kristjansson JK, Garrett R, Zillig W (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. Mol Microbiol 34:217–226

Arnold HP, Zillig W, Ziese U, Holz I, Crosby M, Utterback T, Weidmann JF, Kristjanson JK, Klenk HP, Nelson KE, Fraser CM (2000) A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. Virology 267:252–266

Banner DW, Kokkinidis M, Tsernoglou D (1987) Structure of the ColE1 rop protein at 1.7 A resolution. J Mol Biol 196:657–675

Bartolucci S, Contursi P, Fiorentino G et al (2013) Responding to toxic compounds: a genomic and functional overview of Archaea. Front Biosci 18:165–189

Bell SD (2005) Archaeal transcriptional regulation–variation on a bacterial theme? Trends Microbiol 13:262–265

Bell SD, Jackson SP (2000) Mechanism of autoregulation by an archaeal transcriptional repressor. J Biol Chem 275:31624–31629

Bell SD, Kosa PL, Sigler PB, Jackson SP (1999) Orientation of the transcription preinitiation complex in archaea. Proc Natl Acad Sci USA 96:13662–13667

Bell SD, Magill CP, Jackson SP (2001) Basal and regulated transcription in Archaea. Biochem Soc Trans 29:392–395

Brennan RG, Matthews BW (1989) The helix-turn-helix DNA binding motif. J Biol Chem 264:1903–1906

Brinkman AB, Bell SD, Lebbink RJ et al (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. J Biol Chem 277:29537–29549

Brinkman AB, Ettema TJ, de Vos WM, van der Oost J (2003) The Lrp family of transcriptional regulators. Mol Microbiol 48:287–294

Brock TD, Brock KM, Belly RT, Weiss RL (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Microbiol 84:54–68

Calvo JM, Matthews RG (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Rev 58:466–490

Cannio R, Fiorentino G, Carpinelli P, Rossi M, Bartolucci S (1996) Cloning and overexpression in *Escherichia coli* of the genes encoding NAD-dependent alcohol dehydrogenase from two *Sulfolobus* species. J Bacteriol 178:301–305

Cannio R, Fiorentino G, Rossi M, Bartolucci S (1999) The alcohol dehydrogenase gene: distribution among *Sulfolobales* and regulation in *Sulfolobus solfataricus*. FEMS Microbiol Lett 170:31–39

Chen HT, Legault P, Glushka J, Omichinski JG, Scott RA (2000) Structure of a (Cys3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. Protein Sci 9:1743–1752

Chen L, Brügger K, Skovgaard M, Redder P, She Q, Torarinsson E, Greve B, Awayez M, Zibat A, Klenk HP, Garrett RA (2005) The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. J Bacteriol 187:4992–4999

Coles M, Djuranovic S, Söding J, Frickey T, Koretke K, Truffault V, Martin J, Lupas AN (2005) AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels. Structure 13:919–928

Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, Bartolucci S (2003) Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. FEMS Microbiol Lett 218:115–120

Contursi P, Jensen S, Aucelli T, Rossi M, Bartolucci S, She Q (2006) Characterization of the *Sulfolobus* host–SSV2 virus interaction. Extremophiles 10:615–627

Contursi P, Cannio R, Prato S, She Q, Rossi M, Bartolucci S (2007) Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. J Bacteriol 189:6339–6350

Contursi P, Cannio R, She Q (2010) Transcription termination in the plasmid/virus hybrid pSSVx from*Sulfolobus islandicus*. Extremophiles 14:453–463

Contursi P, D'Ambrosio K, Pirone L, Pedone E, Aucelli T, She Q, De Simone G, Bartolucci S (2011) C68 from the *Sulfolobus* islandicus plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. Biochem J 435:157–166

Di Fiore A, Fiorentino G, Vitale RM et al (2009) Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in Archaea by MarR family proteins. J Mol Biol 388:559–569

Duan X, He ZG (2011) Characterization of the specific interaction between archaeal FHA domain-containing protein and the promoter of a flagellar-like gene-cluster and its regulation by phosphorylation. Biochem Biophys Res Commun 407:242–247

Eberle W, Pastore A, Sander C, Rosch P (1991) The structure of ColE1 rop in solution. J Biomol NMR 1:71–82

Enoru-Eta J, Gigot D, Thia-Toong TL et al (2000) Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius*

homologous to the bacterial global transcriptional regulator Lrp. J Bacteriol 182:3661–3672
Enoru-Eta J, Gigot D, Glansdorff N, Charlier D (2002) High resolution contact probing of the Lrp-like DNA-binding protein Ss-Lrp from the hyperthermoacidophilic crenarchaeote *Sulfolobus solfataricus* P2. Mol Microbiol 45:1541–1555
Ettema TJ, Brinkman AB, Tani TH, Rafferty JB, Van Der Oost J (2002) A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. J Biol Chem 277:37464–37468
Ettema TJ, Huynen MA, de Vos WM, van der Oost J (2003) TRASH: a novel metal-binding domain predicted to be involved in heavy-metal sensing, trafficking and resistance. Trends Biochem Sci 28:170–173
Ettema TJ, Brinkman AB, Lamers PP et al (2006) Molecular characterization of a conserved archaeal copper resistance (cop) gene cluster and its copper-responsive regulator in *Sulfolobus solfataricus* P2. Microbiology 152:1969–1979
Fiorentino G, Cannio R, Rossi M, Bartolucci S (2003) Transcriptional regulation of the gene encoding an alcohol dehydrogenase in the archaeon *Sulfolobus solfataricus* involves multiple factors and control elements. J Bacteriol 185:3926–3934
Fiorentino G, Ronca R, Cannio R et al (2007) MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*. J Bacteriol 189:7351–7360
Fiorentino G, Ronca R, Bartolucci S (2009) A novel *E. coli* biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein. Appl Microbiol Biotechnol 82:67–77
Fiorentino G, Del Giudice I, Bartolucci S et al (2011) Identification and physicochemical characterization of BldR2 from *Sulfolobus solfataricus*, a novel archaeal member of the MarR transcription factor family. Biochemistry 50:6607–6621
Fusco S, She Q, Bartolucci S, Contursi P (2013) T(lys), a newly identified *Sulfolobus* spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. J Virol 87:5926–5936
Gajiwala KS, Burley SK (2000) Winged helix proteins. Curr Opin Struct Biol 10:110–116
Geiduschek EP, Ouhammouch M (2005) Archaeal transcription and its regulators. Mol Microbiol 56:1397–1407
Guillière F, Peixeiro N, Kessler A, Raynal B, Desnoues N, Keller J, Delepierre M, Prangishvili D, Sezonov G, Guijarro JI (2009) Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. J Biol Chem 284:22222–22237
Guillière F, Danioux C, Jaubert C, Desnoues N, Delepierre M, Prangishvili D, Sezonov G, Guijarro JI (2013) Solution structure of an archaeal DNA binding protein with an eukaryotic zinc finger fold. PLoS ONE 8:e52908–0052908
Guo L, Brügger K, Liu C, Shah SA, Zheng H, Zhu Y, Wang S, Lillestøl RK, Chen L, Frank J, Prangishvili D, Paulin L, She Q, Huang L, Garrett RA (2011) Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. J Bacteriol 193:1672–1680
Helmer-Citterich M, Anceschi MM, Banner DW, Cesareni G (1988) Control of ColE1 replication: low affinity specific binding of Rop (Rom) to RNAI and RNAII. EMBO J 7:557–566
Hsu CH, Wang AH (2011) The DNA-recognition fold of Sso7c4 suggests a new member of SpoVT-AbrB superfamily from archaea. Nucleic Acids Res 39:6764–6774
Huffman JL, Brennan RG (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. Curr Opin Struct Biol 12:98–106
Jaubert C, Danioux C, Oberto J, Cortez D, Bize A, Krupovic M, She Q, Forterre P, Prangishvili D, Sezonov G (2013) Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. Open Biol 3:130010
Keeling PJ, Klenk HP, Singh RK, Feeley O, Schleper C, Zillig W, Doolittle WF, Sensen CW (1996) Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. Plasmid 35:141–144
Kessler A, Sezonov G, Guijarro JI, Desnoues N, Rose T, Delepierre M, Bell SD, Prangishvili D (2006) A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. Nucleic Acids Res 34:4837–4845
Kraft P, Kümmel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM (2004a) Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol 78:7438–7442
Kraft P, Oeckinghaus A, Kümmel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM (2004b) Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. J Virol 78:11544–11550
Krishna SS, Majumdar I, Grishin NV (2003) Structural classification of zinc fingers: survey and summary. Nucleic Acids Res 31:532–550
Kumarevel T, Nakano N, Ponnuraj K et al (2008a) Crystal structure of glutamine receptor protein from *Sulfolobus tokodaii* strain 7 in complex with its effector L-glutamine: implications of effector binding in molecular association and DNA binding. Nucleic Acids Res 36:4808–4820
Kumarevel T, Tanaka T, Nishio M et al (2008b) Crystal structure of the MarR family regulatory protein, ST1710, from *Sulfolobus tokodaii* strain 7. J Struct Biol 161:9–17
Kumarevel T, Tanaka T, Umehara T, Yokoyama S (2009) ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. Nucleic Acids Res 14:4723–4735
Larson ET, Eilers B, Menon S, Reiter D, Ortmann A, Young MJ, Lawrence CM (2007a) A winged-helix protein from *Sulfolobus* turreted icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus. Virology 368:249–261
Larson ET, Eilers BJ, Reiter D, Ortmann AC, Young MJ, Lawrence CM (2007b) A new DNA binding protein highly conserved in diverse crenarchaeal viruses. Virology 363:387–396
Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, Douglas T, Young MJ (2009) Structural and functional studies of archaeal viruses. J Biol Chem 284:12599–12603
Lipps G (2006) Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus*. Extremophiles 10:17–28
Lipps G, Ibanez P, Stroessenreuther T, Hekimian K, Krauss G (2001a) The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. Nucleic Acids Res 29:4973–4982
Lipps G, Stegert M, Krauss G (2001b) Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRN1, a putative archael plasmid copy control protein. Nucleic Acids Res 29:904–913
Liu J, Lou Y, Yokota H, Adams PD, Kim R, Kim SH (2005) Crystal structure of a PhoU protein homologue: a new class of metalloprotein containing multinuclear iron clusters. J Biol Chem 280:15960–15966
Malgieri G, Russo L, Esposito S, Baglivo I, Zaccaro L, Pedone EM, Di Blasio B, Isernia C, Pedone PV, Fattorusso R (2007) The prokaryotic Cys2His2 zinc-finger adopts a novel fold as revealed

by the NMR structure of *Agrobacterium tumefaciens* Ros DNA-binding domain. Proc Natl Acad Sci USA 104:17341–17346

Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W (1984) SAV1, a temperate UV-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J 3:2165–2168

Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzer E, Young MJ, Bothner B, Lawrence CM (2008) Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology 376:270–278

Miyazono K, Tsujimura M, Kawarabayasi Y, Tanokura M (2007) Crystal structure of an archaeal homologue of multidrug resistance repressor protein, EmrR, from hyperthermophilic archaea *Sulfolobus tokodaii* strain 7. Proteins 67:1138–1146

Nakano N, Kumarevel T, Matsunaga E, Shinkai A, Kuramitsu S, Yokoyama S (2007) Purification, crystallization and preliminary X-ray crystallographic analysis of ST1022, a putative member of the Lrp/AsnC family of transcriptional regulators isolated from *Sulfolobus tokodaii* strain 7. Acta Crystallogr Sect F Struct Biol Cryst Commun 63:964–966

Napoli A, van der Oost J, Sensen CW et al (1999) An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. J Bacteriol 181:1474–1480

Nguyen-Duc T, Peeters E, Muyldermans S, Charlier D, Hassanzadeh-Ghassabeh G (2013) Nanobody(R)-based chromatin immunoprecipitation/micro-array analysis for genome-wide identification of transcription factor DNA binding sites. Nucleic Acids Res 41:e59

Oke M, Carter LG, Johnson KA, Liu H, McMahon SA, Yan X, Kerou M, Weikart ND, Kadi N, Sheikh MA, Schmelz S, Dorward M, Zawadzki M, Cozens C, Falconer H, Powers H, Overton IM, van Niekerk CA, Peng X, Patel P, Garrett RA, Prangishvili D, Botting CH, Coote PJ, Dryden DT, Barton GJ, Schwarz-Linek U, Challis GL, Taylor GL, White MF, Naismith JH (2010) The Scottish structural proteomics facility: targets, methods and outputs. J Struct Funct Genomics 11:167–180

Oppermann UCT, Knapp S, Sonetto V, Ladenstein R, Jörnvall H (1998) Isolation and structure of repressor-like proteins from the archaeon *Sulfolobus solfataricus*: Co-purification of RNase A with Sso7c. FEBS Lett 432:141–144

Palm P, Schleper C, Gramppa B, Yeats S, McWilliam P, Reiter WD, Zillig W (1991) Complete nucleotide sequence of the virus SSV1 of *the archaebacterium Sulfolobus shibatae*. Virology 185:242–250

Pedone E, Bartolucci S, Fiorentino G (2004) Sensing and adapting to environmental stress: the archaeal tactic. Front Biosci. 9:2909–26

Peeters E, Charlier D (2010) The Lrp family of transcription regulators in archaea. Archaea 2010:750457

Peeters E, Thia-Toong TL, Gigot D, Maes D, Charlier D (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. Mol Microbiol 54:321–336

Peeters E, Albers SV, Vassart A et al (2009) Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase-encoding operon and permease genes. Mol Microbiol 71:972–988

Peeters E, Peixeiro N, Sezonov G (2013) Cis-regulatory logic in archaeal transcription. Biochem Soc Trans 41:326–331

Peixeiro N, Keller J, Collinet B, Leulliot N, Campanacci V, Cortez D, Cambillau C, Nitta KR, Vincentelli R, Forterre P, Prangishvili D, Sezonov G, Tilbeurgh H (2013) Structure and function of AvtR, a novel transcriptional regulator from a hyperthermophilic archaeal lipothrixvirus. J Virol 87:124–136

Peng X, Blum H, She Q, Mallok S, Brügger K, Garrett RA, Zillig W, Prangishvili D (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology 291:226–234

Peng X, Garrett RA, She Q (2012) Archaeal viruses—novel, diverse and enigmatic. Sci China Life Sci 55:422–433

Perera IC, Grove A (2010) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. J Mol Cell Biol 2:243–254

Pérez-Rueda E, Janga SC (2010) Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin. Mol Biol Evol 27:1449–1459

Pina M, Bize A, Forterre P, Prangishvili D (2011) The archeoviruses. FEMS Microbiol Rev 35:1035–1054

Prangishvili D, Arnold HP, Gotz D, Ziese U, Holz I, Kristjansson JK, Zillig W (1999) A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. Genetics 152:1387–1396

Prangishvili D, Stedman K, Zillig W (2001) Viruses of the extremely thermophilic archaeon *Sulfolobus*. Trends Microbiol 9:39–43

Prangishvili D, Garrett RA, Koonin EV (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res 117:52–67

Prato S, Cannio R, Klenk HP, Contursi P, Rossi M, Bartolucci S (2006) pIT3, a cryptic plasmid isolated from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* IT3. Plasmid 56:35–45

Prato S, Vitale RM, Contursi P, Lipps G, Saviano M, Rossi M, Bartolucci S (2008) Molecular modeling and functional characterization of the monomeric primase-polymerase domain from the *Sulfolobus solfataricus* plasmid pIT3. FEBS J 275:4389–4402

Ramos JL, Martínez-Bueno M, Molina-Henares AJ et al (2005) The TetR family of transcriptional repressors. Microbiol Mol Biol Rev 69:326–356

Redder P, Peng X, Brügger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ Microbiol 11:2849–2862

Reiter WD, Palm P, Yeats S, Zillig W (1987) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol Gen Genet 209:270–275

Reiter WD, Palm P, Zillig W (1988) Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res 16:1–19

Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ (2009) Biogeography of the *Sulfolobus islandicus* pan-genome. Proc Natl Acad Sci USA 106:8605–8610

Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S, McDermott T, Young MJ (2001) Viruses from extreme thermal environments. Proc Natl Acad Sci USA 98:13341–13345

Rice G, Tang L, Stedman K, Roberto F, Spuhler J, Gillitzer E, Johnson JE, Douglas T, Young M (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc Natl Acad Sci USA 101:7716–7720

Rigali S, Derouaux A, Giannotta F, Dusart J (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem 277:12507–12515

Sauer RT, Yocum RR, Doolittle RF et al (1982) Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature 298:447–451

Schelert J, Dixit V, Hoang V et al (2004) Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. J Bacteriol 186:427–437

Schelert J, Drozda M, Dixit V et al (2006) Regulation of mercury resistance in the crenarchaeote *Sulfolobus solfataricus*. J Bacteriol 188:7141–7150

Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakić M, Young MJ, Lawrence CM, Copié V (2012) Structural studies of E73 from a hyperthermophilic archaeal virus identify the "RH3" domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. Biochemistry 51:2899–2910

Schreiter ER, Drennan CL (2007) Ribbon-helix-helix transcription factors: variations on a theme. Nat Rev Microbiol 5:710–720

She Q, Singh RK, Confalonieri F et al (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc Natl Acad Sci USA 98:7835–7840

Song N, Nguyen Duc T, van Oeffelen L et al (2013) Expanded target and cofactor repertoire for the transcriptional activator LysM from *Sulfolobus*. Nucleic Acids Res 41:2932–2949

Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res Microbiol 154:295–302

Tenorio-Salgado S, Huerta-Saquero A, Perez-Rueda E (2011) New insights on gene regulation in archaea. Comput Biol Chem 35:341–346

Vassart A, Van Wolferen M, Orell A et al (2013) Sa-Lrp from *Sulfolobus acidocaldarius* is a versatile, glutamine-responsive, and architectural transcriptional regulator. Microbiology 2:75–93

Villafane AA, Voskoboynik Y, Cuebas M et al (2009) Response to excess copper in the hyperthermophile *Sulfolobus solfataricus* strain 98/2. Biochem Biophys Res Commun 385:67–71

Wang B, Yang S, Zhang L, He ZG (2010) Archaeal eukaryote-like serine/threonine protein kinase interacts with and phosphorylates a forkhead-associated-domain-containing protein. J Bacteriol 192:1956–1964

Weininger U, Zeeb M, Neumann P, Löw C, Stubbs MT, Lipps G, Balbach J (2009) Structure-based stability analysis of an extremely stable dimeric DNA binding protein from *Sulfolobus islandicus*. Biochemistry 48:10030–10037

Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M (2004) Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol 78:1954–1961

Wilce JA, Vivian JP, Hastings AF, Otting G, Folmer RH, Duggin IG, Wake RG, Wilce MC (2001) Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. Nat Struct Biol 8:206–210

Wojtas MN, Abrescia NG (2013) Archaeal transcription: making up for lost time. Biochem Soc Trans 41:356–361

Wu RY, Zhang RG, Zagnitko O, Dementieva I, Maltzev N, Watson JD, Laskowski R, Gornicki P, Joachimiak A (2003) Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. J Biol Chem 278:20240–20244

Xiang X, Chen L, Huang X, Luo Y, She Q, Huang L (2005) *Sulfolobus tengchongensis* spindle-shaped virus STSV1: virus-host interactions and genomic features. J Virol 79:8677–8686

Yokoyama K, Nogami H, Kabasawa M et al (2009) The DNA-recognition mode shared by archaeal feast/famine-regulatory proteins revealed by the DNA-binding specificities of TvFL3, FL10, FL11 and Ss-LrpB. Nucleic Acids Res 37:4407–4419